

## **Collaborative Research: North East Cyberinfrastructure Consortium Full Project Detail with Rhode Island Specific Activity**

Reporting period – September 1, 2010 – August 31, 2011

Due: June 1, 2011 (include projected activities for June, July and August in report)

### Report Sections:

#### Project Summary -

The North East Cyberinfrastructure Consortium is in its second year of Track-2 funding. In the second annual meeting held on March 15 at the University of Maine, Orono, we reviewed progress on the installation of fiber for the regional network, cyber-enabled research, workforce development and diversity, external engagement and sustainability. Dr. Jennifer Schopf 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.

#### What we were trying to accomplish:

- 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 manor; pilot projects to foster collaborations across the region;
- Establish collaborations on cybereducation, cyber-knowledgeable workforce development and diversity and outreach;
- Leverage resources to accomplish our goals;

#### Our progress:

- The fiber installation and upgrade projects are in various stages; some have been completed;
- We are carrying out pilot 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 projects inspire novel collaborations on water and cyber-enabled research in the NECC region;
- We collaborate on the Watershed Project to leverage the water education outreach programs in the Region for workforce development and diversity;
- NECC fosters novel collaborations that would not have happened without NECC and its cyber-infrastructure;

- We have examples of small businesses retained because of the new fiber and promise of fiber;

A sample of novel aspects:

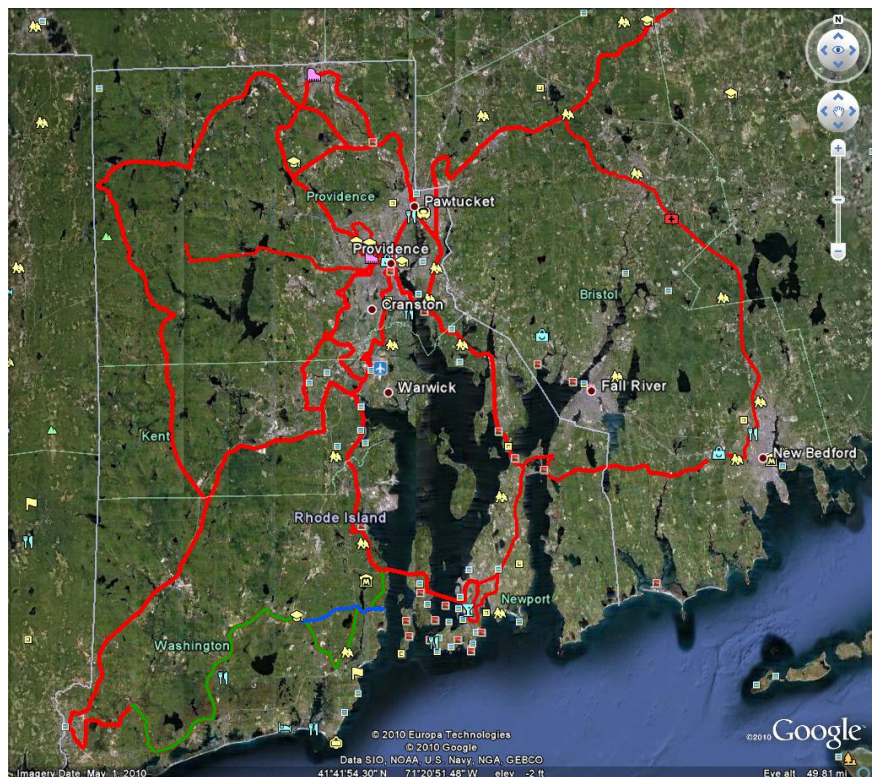
Cyber-enabled research collaborations that require NECC resources.

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

Rhode Island Project Overview:

Rhode Island has made significant progress in its cyberinfrastructure through partnerships and collaborations among Rhode Island organizations. Roughly \$40 M of funding came in through several mechanisms to better connect the state including the EPSCoR RII Track-2 Award. The collaborative project will lay 339 miles of fiber optic cable to connect almost every community in the state- including 50 schools and government offices.

In collaboration with Ocean State Higher Education Assistance Network (OSHEAN), who also collaborated with the Ocean State Libraries and Broadband Rhode Island, a cyber map was planned and URI agreed to lay the fiber for 40 miles in southern portion of the state. The map indicates where OSHEAN is building fiber in red and whereas URI is building fiber in green and blue. The University of Rhode Island staff in Information Technology Services spent months brokering the agreement with Cox



Communications. Building began in May 2011 and 80% contractually must be completed by July 1, 2011. URI also received grant awards from the National Institutes of Health IDeA

Networks of Biomedical Research Excellence (INBRE) program and U.S. Department of Energy (DoE) to fund this endeavor.



Rhode Island also hired a full time Information Technologist to support the website management of the NECC ([www.necyberconsortium.org](http://www.necyberconsortium.org)). Kate Wilson has revamped the existing website into a collaborative format to allow consortium members to manage content.

Kate Wilson also collaborated with Steve Exler from Vermont to provide a web based application for RACER proposals.

Collaboration and outreach were strengthened through the Track-2 grant, Rhode Island EPSCoR was able to leverage its Track-1 grant to support high school and undergraduate students to participate in the stream ecology.

Participation by a Rhode Island Researcher in the Skate Genome projects strengthened the communications and collaborations between the member states.

## **Intellectual Merit -**

### **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.

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. We are now working on generating a new assembly using all of the new data.

#### 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 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 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.

#### **Broader Impacts -**

##### Little Skate Annotation Workshops and Jamborees

The second Skate Genome Annotation Workshop was held at Mount Desert Island Biological Laboratory (MDIBL) on October 12-15, 2010. Over four days, eight instructors



trained 17 participants representing all five NECC states to find and annotate transcriptome and genome sequences for genes in two different pathways and one gene family. The workshop was organized to predominantly have hands-on time working on specific genes rather than the didactic approach of the first workshop. Participants annotated 45 genes from the hedgehog signaling pathway, 54 genes from the WNT signaling pathway and 40 ABC transporters. Interspersed with the workshop time were presentations about gene and protein annotation and others that described each pathway and the ABC transporter gene family. In addition, there were two presentations about lessons learned from previous genome and transcriptome annotation projects. Prof. James Coffman (MDIBL) spoke about his role as an



annotation group leader in the purple sea urchin (*S. purpuratus*) genome annotation project. Prof. Carol Bult (Jackson Laboratory) spoke about her role as a leader in annotating several microbial genomes along with her more recent work annotating the mouse transcriptome. Among the eight participants from Maine, we had four from MDIBL, two from The University of Maine at Orono, including a faculty member, and faculty members from the University of Maine at Machias and the University of Southern Maine. We had a faculty member and graduate student attend from University of New Hampshire. From Vermont, we had a faculty member from Norwich University, a graduate student from the University of Vermont and a student from Johnson State College attend. A faculty member from Rhode Island College and a graduate student and post-doctoral fellow from University of Rhode Island attended. A graduate student from Delaware State University also attended. MDIBL wrote a press release about the workshop

([http://www.mdibl.org/mdibl\\_press\\_releases/Skate\\_Genome\\_Workshop\\_Held\\_at\\_MDIBL/288/](http://www.mdibl.org/mdibl_press_releases/Skate_Genome_Workshop_Held_at_MDIBL/288/)) and a newspaper article was published about the workshop in the Bar Harbor Times (<http://mdi.villagesoup.com/place/story/skate-genome-workshop-held-at-mount-desert-island-biological-laboratory/358443>).

Three genome workshop and annotation jamborees were held in Delaware and Maine in support of the Little Skate genome-sequencing project. The goal of these four to five day workshops was expanding the base of bioinformatics-savvy researchers across the region to catalyze high-impact cyber-enabled research. Each of the workshops included participants from all five NECC states and focused on various aspects of the genome sequencing and annotation. The initial workshop (University of Delaware, May 2010) provided over thirty participants with extensive background knowledge in genomic biology, sequencing technology, and bioinformatics, and allowed them to apply this

knowledge in hand-on exercises. The second workshop (Mount Desert Island Biological Laboratory ME, October 2010) included additional lectures on relevant topics, followed by exercises, where the eighteen participants annotated genes from the little skate transcriptome associated with the Wnt and Hedgehog signaling pathways and ABC transporters. With over forty participants, the final workshop (University of Delaware, May 2011) built upon the results of the earlier workshops with teams focusing on more complete annotation of genomic and transcriptomic genes from the Hedgehog pathway, as well as protein level annotation for a number of targets. This workshop culminated with teams presenting their annotations to the larger group and submitting appropriate information to public annotation repositories.



First Little Skate Genome Annotation Workshop, UD, May 2010

### 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 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 residing on a repurposed six-node cluster provided by the University of Delaware. This BLAST cluster

supported annotation activities of the third skate workshop and will continue to evolve with this and future projects.

#### Programmatic Terms and Conditions

Rhode Island does not have Rhode Island-specific Programmatic Terms and Conditions (PTC).

#### Broadening Participation in Rhode Island

	Men	Women	Minorities
Professional	0	1	
High School	0	3	2
Undergrad	0	1	

#### Increasing Institutional Engagement in Rhode Island

Rhode Island NSF EPSCoR has put effort into promoting water research regarding stream ecology and metagenomics through collaboration with other NECC states. Professor Art Gold from URI's Watershed Watch has collaborated through the Vermont Streams Training. In the summer of 2011, two undergraduate students from Art Gold's laboratory will attend training with Vermont Streams. The undergraduates are Elizabeth Elmstrom and Molly Welsh who are working on the role of beaver pond ecosystems on the flux of greenhouse gases, watershed nitrogen and carbon transformations, and coastal and marine water quality. Associate Professor Rebeka Merson from Rhode Island College is involved with the INBRE side and has participated in the Skate Genome Project and the annotation jamborees held in Delaware.

#### Leveraging

OSHEAN received a \$21.3M grant through the Broadband Technology Opportunities Program (BTOP) at the National Telecommunications and Information Administration (NTIA within the USDOC). As a result of the leveraging of multiple grants, Rhode Island will build a 339-mile fiber optic network to include every county and neighboring Bristol country, Massachusetts. It will connect to over 50 community anchor institutions with a 400Gbps connection that is 50 times faster than those in the homes in Rhode Island. This cyberinfrastructure will also connect to the existing fiber in Boston and Connecticut to connect the other NECC states into the fiber ring. In 2009, the University of Rhode Island received an EPSCoR RII Track-2 award (NSF Award #0918061; \$1.2M). URI also received awards from the National Institutes of Health IDeA Networks of Biomedical Research Excellence (INBRE, (\$750K)) program and the U.S. Department of Energy (\$924K). In 2010, Brown University received NSF EPSCoR's RII Inter-Campus and Intra-Campus Cyber Connectivity grant (NSF Award #1005789).

## **Progress on Program Elements– Year (2)**

*Status* – current status of the consortium’s cyberinfrastructure landscape (in Year 2)

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 is close to completion to 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 almost complete. Details of these networks will be given in the individual state progress reports.

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. To date, NSF EPSCoR has funded \$8.3M in cyberinfrastructure improvements to NECC states.

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					12.3	

## **Leveraged opportunities and activities**

The NECC is highly leveraged. 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.

## **Rhode Island-specific leveraging**

Rhode Island NSF EPSCoR was able to leverage both the cyberinfrastructure funding of other Rhode Island organizations and its own Track-1 funding to broaden its impact. As mentioned, URI collaborated with OSHEAN to plan a multiple loop network throughout the state. In a state that is just 37 miles wide and 48 miles long, there will be 339 miles of fiber connecting over 50 schools, hospitals, and other non-profit entities. To learn more about the Broadband leveraging go to <http://oshean.org/Page/View/btop/Brochure>.

Rhode Island NSF EPSCoR’s Track-1 grant supports a 10-week undergraduate research fellowship for 25 students each summer. In 2010, one of the fellowships was dedicated to water research through the EPSCoR Track-2 Cyberinfrastructure program. The intern was



sent to Vermont for training in stream ecology. Track-1 also supports high school outreach and through Track-2 Rhode Island was able to send a Central Falls high school teacher and three students (minority young women) to the training as well.

*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 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.

Workforce development progress includes a new full time position created in the 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 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.

Table 1: Metagenome Project Collaborators

Water Sample Collection

Sebasticook Lake, ME	Kristin Ditzler, Jasmine Saros	University of Maine
Yawgoo & Trustom Ponds, RI	Linda Green	University of Rhode Island
Lake Winnepesaukee, NH	Jeff Schloss, Jeff Haney	University of New Hampshire
Highgate Springs, VT	Pat Pearson	University of Vermont

## Laboratory Processing

Tim Hunter	University of Vermont
Scott Tighe	University of Vermont
Pat Pearson	University of Vermont

## DNA Sequencing

Bruce Kingham	University of Delaware
---------------	------------------------

## Bioinformatics and Data Management

James Vincent	University of Vermont
Marc Farnum Rendino	University of Vermont and University of Delaware
Chelsea Mitchell	University of Vermont
Colin Delaney	Saint Michael's College, VT
Shawn Polson	University of Delaware

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).

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.

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.

#### Pilot 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 2 jurisdictions. There is no review of pre-proposals. Merit review of proposals is conducted by a subset of the Water Research committee which one representative from each jurisdiction. Vermont EPSCoR administers the proposal evaluation process.

In 2010, there were three pre-proposals and three full RACER proposals; a single award was made to a collaborative effort between investigators at UNH and UVM. Dr. W. Kelley Thomas, UNH Hubbard Professor of Genomics and Director of Genomics Studies, leads the project, "An open-access, shared database resource for eukaryotic metagenetics research." Also on the team are Dr. James Vincent (Bioinformatics Core Director of the Vermont Genetics Network, Vermont INBRE) and Dr. Holly Bik (UNH). 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.

The 2011 RACER process is currently underway with six pre-proposals. Full proposals are due 15 July 2011 with awards to be made in August.

### 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

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 is modeled after the Vermont EPSCoR Streams Project, which engages high school teams and undergraduates in watershed research. In its pilot 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.

- The Genome Annotation workshop is intended to inspire a new generation of bioinformaticians. The workshop attracted 35 attendees to UD in May 2010 – 19 males and one female, on 41 attendees on May 2011, with 29 males and 12 females. The next workshops will have surveys of the attendees.

### *Dissemination and Communication -*

The second year 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 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
- 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

Kate Wilson took over as the information technologist for Rhode Island EPSCoR in January of 2011. Her predecessor, as per the duty of Rhode Island EPSCoR office, hosted the site and maintained a wiki for users to edit the content. The server that hosted the site was unavailable upon hiring so she designed a new website based on the same criteria in the award. The site is now hosted on a server within the 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 former site was archived and taken down.

This site has some pages that are only editable by Kate such as the homepage, the how to page, and partners page. The partners' 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 log in to the edit the pages. Anyone can also comment on any of the posts on the website. Logins can be requested through email with Kate Wilson. As of now, logins exist for Kate, Steve Exler, the Maine EPSCoR office, and Jennifer Specker.

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. 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



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.

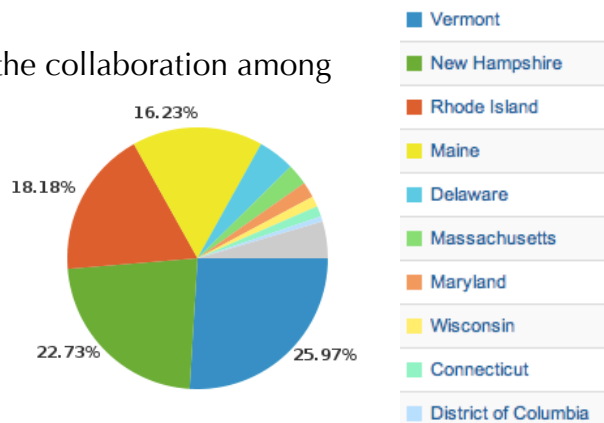
#### Recent Changes

Meetings  
2011-03-22T13:13:09+00:00



UVM hosted application. This page also is imbedded with the NECC site as opposed to taking the visitor from the site.

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 as minutes and reports. The wiki can be enabled to require a login to see certain pages if everything should not be public.



Google Analytics was also added to the site to monitor the number of page hits and where those visitors are coming from. Since March 27<sup>th</sup> (the day it was installed), there have been 158 visits viewing 322 pages. 60% of the views are the homepage. 17% of total page views are the Solicitations page and the rest of the pages share 5% each. 26% of visitors are in Vermont, 23% are from New Hampshire, 18% are from Rhode Island, and 16% are from Maine. The rest are distributed throughout the country.

#### RACER Awards submitted on the website:

- Regional Awards for Cyberenabled Research (RACER) awards were posted in Year 2 and are currently underway; all application and review materials are web based - <http://necyberconsortium.org>
- RACER awardees for Year 1 W. Kelly Thomas - UNH - Meiofaunal metagenetics was awarded in collaboration with Jim Vincent, UVM

#### Communication among NECC Partners:

- Regular NECC videoconferences of Executive Steering Committee and all other committees scheduled each month
- 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. The conference agenda, list of participants, and minutes are attached in Appendix 1 of this report.

#### Planned for Year 2-3:

- Work is underway to produce 3 podcasts and one 30-minute television program made in partnership with Vermont Public Television and Maine Public Broadcasting. Programming will air sometime in Year 3.

#### *Evaluation and Assessment Plan -*

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 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 will occur in May 2011.
- Rhode Island: INBRE supplement assessment occurred in 2010; the next review will occur in 2012.
- Vermont: the AAAS panel in June 2011 will provide formative assessment of the entire NECC program. This is part of the Vermont INBRE. In addition, Vermont EPSCoR is working with Dr. Joy Livingston, Flint Springs Associates, to perform formative assessment for participants in the Annotation Jamborees.

#### *Sustainability Plan -*

At the annual meeting, 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 over the summer of 2011.

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

#### *Management and Coordination Plan -*

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; approved year 2 RFP. The 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, 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;

The other standing committees meet regularly, including the Technical Committee, Water Research Committee and NE Bioinformatics Collaborative. The Technical Committee is 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 the RFP for awards in Year 3 have been disseminated.

#### Unobligated Funds

Rhode Island has \$130,189.22 unobligated funds—less than 20%--as of closing in May 2011.

## **APPENDIX 1:**

### **NECC Year 2 Annual Meeting**

**Hosted by Maine EPSCoR  
March 15, 2011  
University of Maine in Orono**

## 2011 Annual NECC Meeting

March 15, 2011

Wells Conference Center, University of Maine, Orono

### Agenda

- 7:30 a.m. Registration check-in and hearty full breakfast**
- 8:00 a.m. Welcome and Introductions:**
- Judith Van Houten, VT EPSCoR PD and INBRE PI
  - Jennifer Schopf, Program Director, NSF EPSCoR
  - Michael Sayre, Deputy Director, NIH Division of Research Infrastructure
- 8:15 a.m. Introduction of the NECC committees and members**
- 8:30 a.m. NECC Cyber-enabled Collaborative Research Programs:  
(10 minute presentations with 5 minutes Q&A)**
- *NECC Metagenome Project:* James Vincent, NEBC and Vermont INBRE Bioinformatics Core
  - *NECC Little Skate Genome Project:* Ben King, Maine INBRE Bioinformatics Core and NEBC
  - *NECC Cyber-enabled genomics research:* Jennifer Specker, Associate Project Director, Rhode Island NSF EPSCoR, and URI Professor of Oceanography.
  - *NECC RACER YR1 Award for “An open-access, shared database resource for eukaryotic metagenetics research”:* Holly Bik, Postdoctoral Researcher, Hubbard Center for Genome Studies, Univ. of New Hampshire, and James Vincent, NEBC and Vermont INBRE Bioinformatics Core
  - *NECC Sequencing resources:* Bruce Kingham, NEBC, Director, UD Sequencing & Genotyping Center, DBI
- 10:00 a.m. Break**
- 10:15 a.m. Workforce Development and Diversity:  
(10 minute presentations with 5 minutes Q&A)**
- *NECC Little Skate Annotation Jamborees:* Shawn Polson, Coordinator of the Bioinformatics Core Facility, Center for Bioinformatics & Computational Biology, U. Delaware
  - *Cybertools:* Karl Steiner, NECC Executive Committee and Sr. Assoc. Provost for Research Development, U. Delaware
  - *NECC Watershed Project:* Miranda Lescaze, Vermont EPSCoR CWDD and Streams Project Director and NECC Watershed Project
- 11:00 a.m. NECC Committee Meetings (breakouts)**
- NEBC
  - Technical Committee
  - Water Research/RACER Committee
  - Executive Committee
  - Watershed Project
- Noon Lunch (provided), networking, poster session**



- 1:00 p.m. NECC State Projects: Fiber, Data Centers, Sustainability, and Leveraging: (10 minute presentations with 5 minutes Q&A)**
- *Rhode Island Fiber Progress and NECC Website:* David Porter, Director of the University of Rhode Island Media and Technology Services, NECC Technical Committee
  - *Vermont Fiber progress:* Kelvin Chu, Assoc. PD, VT EPSCoR, NECC Executive Committee
  - *New Hampshire Fiber Progress:* Scott Valcourt, Director, Project Management and Consulting Services, UNH and NECC Executive and Technical Committees
  - *Delaware Progress and Data Center:* Karl Steiner, NECC Executive Committee, Sr. Assoc. Provost for Research Development
  - *Maine Fiber progress and Data Center:* NECC Technical Committee Co-Chairs Jeff Letourneau, Executive Director, NetworkMaine, University of Maine System, and Bruce Segee, Butler Associate Professor of Electrical and Computer Engineering, University of Maine
  - *NECC Data Management:* Ben King, Maine INBRE Bioinformatics Core, NEBC
- 2:45 p.m. Break**
- 3:00 p.m. NECC Evaluation and Assessment:**
- *Special Report from the Executive Committee:* Kelvin Chu, Assoc. PD VT EPSCoR, NECC Executive Committee
- 3:30 p.m. NECC External Engagement:**
- *Communication, Outreach, and Dissemination Partnerships:* Lillian Gamache, Project Administrator, VT EPSCoR, and Vicki Nemeth, Assoc. PD, ME EPSCoR Director, and NECC Executive Committee
- 4:00 p.m. NECC Committee reports and updates:**
- NEBC
  - Technical Committee
  - Water Research/RACER Committee
  - Executive Committee
- 5:00 p.m. Planning ahead:**
- Annual report & assessment discussion
  - Next Track-2
  - What can we do now that we could not do before?
  - Moving forward with additional collaborative research projects
- 6:00 p.m. Dinner on own**

2011 Annual NECC Meeting Attendees  
March 15, 2011

Last Name	First Name	Title/Position	Institution/Organization
Bik	Holly	Postdoc	University of New Hampshire
Broadhurst	Amy	Assistant Director/Delaware EPSCoR PA	University of Delaware, Delaware Environmental Institute
Chu	Kelvin	VT EPSCoR Associate Project Director	University of Vermont
Cousins	Stephen	Supercomputer Engineer/Administrator	University of Maine
Eckardt	Michael	Vice President for Research	University of Maine
Gamache	Lillian	VT EPSCoR Project Coordinator	University of Vermont
Green	Douglas	Network Architect	University of New Hampshire
Gregoire	Michelle	Program Manager	New Hampshire EPSCoR
Gregory	John	Executive Director, Information Technologies	University of Maine.Maine INBRE
Huang	Xiuzhen	Associate Professor	Arkansas State University
King	Benjamin	Staff Scientist	Mount Desert Island Biological Laboratory
Koskie	John	Operations Manager	University of Maine
Lescaze	Miranda	VT EPSCoR CWDD and Streams Project Director	University of Vermont
Letourneau	Jeff	Executive Director, Networkmaine	University of Maine System
Mattingly	Carolyn	Associate Professor	Mount Desert Island Biological Laboratory
Nemeth	Vicki	Director of Research Administration & Maine EPSCoR	University of Maine
Parr	Thomas	Sustainability Science Initiative Graduate Student	University of Maine
Peckenham	John	Director, Maine Water Resource Research Institute	Mitchell Center, University of Maine

Polson	Shawn	Assistant Professor/Bioinformatics Core Coordinator	University of Delaware
Porter	David	Director, Media & Technology Services	University of Rhode Island
Saros	Jasmine	Associate Professor	University of Maine
Schloss	Jeffrey	Extension Professor in Biology and Water Quality Specialist	University of New Hampshire Cooperative Extension
Segee	Bruce	Technical Director, Maine Supercomputer	University of Maine
Shaikh	Zahir	Director, RI-INBRE	University of Rhode Island
Specker	Jennifer	Associate Project Director, Rhode Island NSF	University of Rhode Island
Steiner	Karl	Senior Associate Provost for Research Development	University of Delaware
Strock	Kristin	Graduate Student	University of Maine
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
Zhu	Yifeng	Associate Professor	University of Maine
Schopf	Jennifer	Program Director	National Science Foundation
Sayre	Michael	Deputy Director	NIH Division of Research Infrastructure (via videoconference)
Taylor	Fred	IDeA Director	NIH (via videoconference)
5 others		Program Officers	NIH (via videoconference)

## **2011 Second Annual NECC Meeting**

**March 15, 2011**

**Wells Conference Center, University of Maine, Orono**

### **Minutes**

#### **A) Welcome and Introductions:**

- Judith Van Houten, VT EPSCoR PD and INBRE PI
  - Where did we start? IBM, State of Vermont, Higher Ed
  - NECC Overview
    - First plans launched among NE INBRE PIs in 2006
    - Face-to-face meeting to discuss fiber projects in Vermont in 2008
    - *Collaborative* NSF EPSCoR Track-2, submitted Jan. 2009: (\$6M)
    - Supplement requests to NIH-NCRR in Spring 2009: \$8.4M
    - NSF – C2 awards to DE & RI in Sept. 2010: \$2.3M
    - Total Awards to Date: IH-NCRR\$8.4M; NSF-EPSCoR \$8.3M
    - NECC annual meeting – 2010 in Burlington, VT; 2011 in Orono, ME
  - 28 October 2010 Press Release
    - Over \$17M in competitive awards from the National Science Foundation (NSF) and National Institutes of Health (NIH) to the Northeast Cyberinfrastructure Consortium (NECC)
  - Pilot awards for Novel Scientific Collaboration
    - 2010: Meiofaunal Metagenetics
    - Dr. W. Kelly Thomas, UNH Hubbard Professor of Genomics, Director of Genomic Studies; Dr. Holly Bik, UNH, Dr. Jim Vincent, UVM
    - Establish a database structure and data repository for eukaryotic metagenetic studies using the NECC data center
- Jennifer Schopf, NSF EPSCoR
- Michael Sayre, Deputy Director, Division of Research Infrastructure

#### **B) Introduction of the NECC committees and members**

#### **C) NECC Cyber-enabled Collaborative Research Programs:**

- *NECC Metagenome Project*: James Vincent, NEBC and Vermont INBRE Bioinformatics Core
  - Metagenomic Survey of Blue-Green Algae: A Pilot Project of the NECC
  - ARRA Pilot Project
  - One of two supplements to VGN INBRE
  - ARRA Supplement Priorities: Create jobs, Build infrastructure, Develop expertise to continue 1 & 2
  - Why and how do cyanobacteria dominate?
  - Metagenomic Approach Part I – Who is there? Compare Samples
  - Metagenomic Approach Part II –What are they doing?

- Sebasticook Lake, Maine: Kristin Ditzler, Jasmine Saros, University of Maine
- Yawgoo Pond & Trustom Pond, RI: Linda Green, University of Rhode Island
- Lake Winnepesaukee, NH: Jeff Scholoss, Jeff Haney, University of New Hampshire
- Highgate Springs, VT: Pat Perason, University of Vermont
- Select Samples: Begin – Peak – End; Bloom Density Over Time
- Sample Collection: 164 samples x 6 preps each = lots of work
- Algae Counting Results
- To Do: Sequencing
- 16's rDNA Survey Horizontal Screen
- Deep Sequencing – Vertical Screen (VT Samples Only)
- Metagenomics, Genomics, Metagenetics, ?
- Regional Collaborative Research Projects
- NEBC, Core Facilities
- BioIT Personnel, Shared Data Center: NECC Regional Network: VT, NH, ME, RI, DE
- ARRA Results: Created one full time job, one temporary, Developed Expertise, Science results to come...
- Funding
  - National Center for Research Resources
    - P20RR16462 Vermont Genetics Network – Vermont INBRE
    - P20RR016463 Comparative Functional Genomics INBRE in Maine
    - P20RR016457 Rhode Island INBRE
    - P20RR018787 Cellular and Molecular Mechanisms of Lung Disease
    - P20 RR016472 Delaware INBRE
  - NSF
    - EPS-0918284 University of Vermont
    - EPS-0918033 University of New Hampshire
    - EPS-0918078 University of Delaware
    - EPS-0918018 University of Maine
    - EPS-0918061 University of Rhode Island
- Who does all this? Grant writing, Experimental design, Sample collection, Sample processing, Sequencing, Project management, Data storage, management, Bioinformatics, Manuscript writing
- Bait and Switch: The North East Cyberinfrastructure Consortium
  - NECC Goals: Optical Network, Personnel Infrastructure, Leverage for research
  - Regional Fiber Optic Network: VT, NH, ME, RI, DE; 2 to 120 Gb for same yearly cost UVM
  - Shared Infrastructure: IT People, Data Center,
- NECC Shared Data Center
  - Shared Data Center (SDC)
  - Cloud Based Services ↔ University of Delaware ↔ University of Maine
  - Shared Data Center File Storage; Shared Data Center Relational Data Base Support



- On Demand hosted MySQL with real people support
  - Resources: NEBC Core Facilities, IT People Shared Data Center, Regional NetworkCore Facilities, Vgn.uvm.edu, Delaware Biotechnology Institute DNA Sequencing Center, - UVM Massively Parallel Sequencing, North East Bioinformatics Collaborative NEBC,
  - The Little Skate Genome Project
    - Skate genome > human genome
  - NEBC Workshops: May 2010, October 2010, May 2011, Next workshop October 12-15, Maine
- *NECC Little Skate Genome Project*: Ben King, Maine INBRE Bioinformatics Core and NEBC
- The Genome Sequencing Project for The Little Skate: Limbed Vertebrates, Limb Regeneration
  - Collaborative Research & Training
    - Skate Reagents (Maine) → Sequencing (Delaware) → Data Storage (Regional Data Center) → Assembly (Vermont, Maine) → Annotation (Delaware, Maine, Vermont, NH, RI) → Public Access (Delaware, Maine) → Research Applications (Maine, Rhode Island)
  - Whole Genome Shotgun Sequencing: Randomly Fragment Genomic DNA, Sequence Fragments, Genome Assembly, Contiguous Sequence (Contig)
  - Status of Genome Sequencing and Assembly
    - Short Fragment Reads: 26x Genome Coverage, 100.8 Gbp, 806 million 125 bp reads
    - Genome Assembly: 3 million contigs, Total length of contigs: 1.6 Gbp, N50: 646 bp, Longest contig: 21 kb
  - Leverage MDIBL's Existing Skate Data for Genome Annotation
    - Analyze Expressed Sequences: Determine which genes are present; Find exon; 173,000 Transcriptome Contigs; 260 Million 76bp reads; 31,000 ESTs; MicroRNAs
  - Genome Sequence
  - Expressed Transcripts
  - Shared Data Center Critical To Project
    - Upload and download files through a web interface
    - Reads 76 GB
    - Assembly: 0.6 GB
      - Discovery of Homeobox (HOX) Cluster Deletion: Primitive Shark- Four Clusters;
    - Skate - Missing HOXC Cluster; Human - Four Clusters; First reported loss of an HOX cluster in any animal
- *NECC RACER YR1 Award for "An open-access, shared database resource for eukaryotic metagenetics research"*: W. Kelly Thomas, UNH Chief Sustainability Officer and UNHSA Director, University of New Hampshire, and James Vincent, NEBC and Vermont INBRE Bioinformatics Core
- Eukaryotic metagenetic: High-throughput biodiversity research (Holy Bik)

- High-throughput biodiversity research
  - Oceanic sediments (covering > 70% of the earth's surface) harbor the vast majority of the world's biodiversity
  - Microscopic eukaryotes (e.g. nematode worms, protists, fungi) are diverse and abundant in these environments
  - The taxonomy and functional role of these species (likely to be significant in marine ecosystems) is not understood.
  - Informed mitigation and remediation require prior knowledge of biodiversity.
- Expense of Taxonomic Surveys
  - Time: Contracts take weeks or months to complete
  - Staff
    - Large projects need teams of taxonomists
    - Scientists would rather be doing research
  - Cost
  - Accuracy?
    - Dependent on skill level of staff
    - Only ~ 100 animals per phylum analyzed
- Diverse marine community → Extract environmental DNA → Amplify 'barcoding gene' → high-throughput sequencing → metagenetic community analysis
- Applications of high-throughput biodiversity research
  - Biological recovery of disturbed habitats
  - Monitoring for potential habitat degradation/biota loss from anthropogenic impacts
  - Baseline surveys for offshore drilling
  - Environmental impacts of farmed salmon
- Meiofaunal Biodiversity in the Gulf of Mexico
- Key Questions
  - How unique are the communities in the Gulf of Mexico?
  - How structured are the communities within the Gulf of Mexico?
  - What has been the effect of Anthropogenic disturbance on these communities?
  - To address these questions, we will utilize cutting edge high-throughput DNA sequencing technologies alongside traditional taxonomic methods of analysis in order to assay the diversity of virtually all eukaryotic organisms from samples collected prior to the April 2010 spill.
- Eukaryotic Database Resources
  - Clear mandate from NSF to link high-throughput sequence data and taxonomy
  - In the absence of a robust reference database (and lack of funding to develop one), this database issue is a substantial barrier for robust evolutionary interpretations of metagenetic data
- NECC RACER Award:
  - Design database resources that meet the needs of the eukaryotic community
  - Leverage tools/resources currently geared towards microbial projects
- Important Challenges
  - Guide Trees

- rRNA data needs to be interpreted in a phylogenetic context
- Phylogenetic placement of short sequences can help you identify taxon sampling problems in the reference dataset that would not be obvious by BLAST searches
- However, these analyses are critically dependant on a robust guide tree
- Metadata Analyses
  - Genbank's Short Read Archive is not accessible
  - MOTUs (Molecular Operational Taxonomic Units) are arbitrary constructions
- Current Analytical Tools
  - Camera
    - Focused on metagenomic analysis of microbial communities (shotgun sequencing of environmental DNA)
  - QIIME
    - Comprehensive pipelines for ecological analysis (rRNA genes), but current database resources limited to microbial taxa (Greengenes, RDP)
    - New UNH/QIIME collaboration to focus on eukaryotes
- NECC database
  - Long-term repository to provide a 'snapshot' of data at time of publication
  - Tools for exploring intragenomic rRNA variation specific to eukaryotes
    - Help derive biology/ecology
  - Web portal to facilitate interaction amongst the eukaryotic community after NESCent meeting
- *NECC Sequencing resources:* Bruce Kingham, NEBC, Director, UD Sequencing & Genotyping Center, DBI

#### **D) Workforce Development and Diversity:**

- *NECC Little Skate Annotation Jamborees:* Shawn Polson, Coordinator of the Bioinformatics Core Facility, Center for Bioinformatics & Computational Biology, U. Delaware
  - Skate Genome Project: Little Skate Genome Sequencing & Annotation
  - Collaborative use of specialized resources & expertise in an integrated process
  - INBRE Administrative Supplement (2009-11)
    - Little Skate (*Leucoraja erinacea*) Clones: MDIBL – Mount Desert Island Biological Lab (ME)
    - Next Generation Sequencing: UD DNA Sequencing & Genotyping Center (DE)
    - Sequence Assembly: Vermont Genetics Network (VT) with ME, RI
    - Sequence Analysis/Annotation: Analysis/annotation pipeline at CBCB (DE), ME, RI, NH, VT
    - Storage & Access of Sequence/Annotation data: shared data center (DE), backup ME
    - Public Dissemination: UD CBCB (DE) MDIL (ME), skatebase.org (VT)
    - Skate Genome Project : NECC Cyber Enabled Research, Building a Base of Cyber-Knowledgeable Researchers

- Sequence-Driven Biology
- Next Gen Sequencing
  - Sequencing no longer specialty tool of Molecular Biologists
  - Permeates all fields of biology
  - Affordability encouraging investigators to utilize
- Bioinformatic Bottleneck
  - Researchers often ill prepared for flood of data
  - Bioinformatics often more costly than sequencing
  - Need pool of trained bioinformatically-savvy researchers
- Skate Annotation Workshops
  - Collaborative & Distributed training
    - Integrated with little skate genome sequencing and annotation
    - Genome annotation workshops: lectures, tutorials, annotation exercises
    - Coupling training helps improve understanding of underlying bioinformatics approaches and methods
    - Participants: Trainees from NECC institutions, regional minority/undergraduate institutions
- Genome Annotation Workshops
  - May 24-28, 2010: UD CBCB (DE)
  - October 12-15, 2010: MDIBL (ME)
  - May 23-27, 2011: UD CBCB (DE)
- First Skate Workshop, University of Delaware (May 24-28, 2010)
  - Focus on learning basic concepts in genome annotation
    - Molecular Biology Primer
    - Illumina Genome Sequencing
    - Genome Sequence Analysis and Tools
    - Protein Sequence Analysis
    - Functional Annotation, Ontologies
    - Literature Mining
  - Reinforced with hands-on exercises
- Second Skate Workshop, Mount Desert Island Biological Laboratory (Oct 12-15, 2010)
  - Focus: Little Stake Transcriptome Annotation
  - Lectures on a range of relevant topics
    - Mitochondrial Gene Annotation
    - Gene Pathways of Interest
    - Annotation of the Purple Sea Urchin
    - Students add manual annotation to transcriptome; Hedgehog pathway, WNT pathway, ABC transporter
- Third Skate Workshop, University of Delaware (May 23-27, 2011)
  - Focus: Annotation of genes/proteins in Hedgehog Signaling Pathway
  - Lectures
    - UniPortKB Sequence Annotation Process

- NCBI Sequence Submission Process (Dr. Kim Pruitt, NCBI)
  - Keynote by Dr. Jason Moor. Dartmouth Medical School
- Exercises
  - Manual Annotation of Hedgehog pathway features
  - Submission of full-length annotated sequences to GenBank
  - Student Poster Session
- CBCB Symposium on Bioinformatics and Systems Biology
  - University of Delaware (May 27, 2011)
  - Highlight Computational and bioinformatically-enabled life science research at the University of Delaware
  - Talks and Posters from among the over 40 affiliated research groups across campus
  - Keynote and Funding Agency Roundtable
- *Cybertools*: Karl Steiner, NECC Executive Committee and Sr. Assoc. Provost for Research Development, U. Delaware
  - NECC Cyber Tools
    - Karl V. Steiner, 2011 Annual NECC Meeting; Orono, ME – March 15, 2011
  - Why Do we Need Cyber Tools?
    - Science has never been locally restricted
      - People who have travelled long distances for new knowledge and to work with colleagues to solve scientific challenges
    - Increasingly scientific data are stored digitally and made available instantaneously
    - Regionally, nationally and globally distributed collaborations have become the norm
    - Successful collaborations within this framework require new models and tools to interact successfully – Cyber Tools
  - What are Cyber Tools?
    - Communication Tools
      - Email, videoconferencing, video-streaming
      - Social networking tools
      - Collaboratories
      - Podcasts, Facebook, Twitter
    - Scientific Tools
      - Convert scientific events/biological information into digitized data
      - Genome sequencing & annotation
      - Environmental monitoring, remote sensing
  - Videoconferencing
    - NECC Videoconferencing at Work, Promise vs. Reality
  - Video Streaming
  - DENIN Podcasts
  - Social Networking Tools (Vivo/Prism/Profiles)
    - Social Networking tool for scientific collaboration



- Sematic Web Application
- Ingests many data types: databases, forms, PubMed
- Locate individuals with needed expertise/interests
- Deployed across UD and other Consortia
- Collaboratory at UD-DBI
  - Statewide Access to Bioimaging Instrumentation through Web-based Collaboratory
  - Field Emission Microscope & Multiphoton Confocal Microscope
- Scientific Collaborations
  - Scientific Cyber Tools
    - Genomic Sequencing & Annotation
    - Distributed Database Servers
    - Environmental Monitoring and Analysis
- Genomic Sequencing at UD
  - Illumina HiSeq 1000 Sequencer Upgrade (NSF EPSCoR Upgrade)
  - Pacific Biosciences PacBio RS Sequencer (2010 NSF MRI Grant)
  - Bruce Kingham, Manager UD Sequencing & Genotyping Lab
- NGS – Illumina HiSeq vs PacBio RS
  - Complementary Capabilities
- Next Generation Sequencing (NGS)
- Skate Genome Project; Bioinformatics Workforce Development
  - Collaborative and distributed training
    - Integrated with research on little skate genome sequencing and annotation
    - Genome annotation workshops (DE, ME, DE): One-week long, semiannual covering lectures, tutorials, annotation exercises
  - Annotation Workshops and Jamborees
    - Delaware – May 2010
    - Maine – October 2010
    - Delaware – May 2011
- Watershed Monitoring
  - Use of GIS for gathering original data from White Clay Creek and its surrounding tributaries and helped to produce a research poster of their findings.
  - Compiled GIS based watershed map atlas with layers of land use, impervious cover, topography, wetlands, and stream water quality
  - Collected measurements of stream flow, precipitation, and water quality to analyze and model impacts of land use and climate change on the White Clay Creek, using a new water monitoring probe
  - Data collected has been loaded into Excel and uploaded to Water Resources Agency's webpage for community view <http://www.ipa.udel.edu/wra/>
  - Podcasts of Environmental Career Day, Sensors and Seminars available to NECC EPSCoR Community ([www.denin.udel.edu/podcasts](http://www.denin.udel.edu/podcasts))
  - Future Plans:

- Currenly in planning stages for hands-on sensor training (Dan Leathers – UD) and use of the water monitoring probe (Gerald Kaufmann –WRA)
- Water Resource Agency
- Wesley College EPSCoR-Supported Computer Applications Laboratory
  - Objectives
    - To develop a GIS and computer modeling applications for lab for instruction and research in environmental planning
    - To develop a multidisciplinary general studies GIS course
    - Bridge fields research with GIS and computer modeling
    - Stream Assessment Coupled with Land-use modeling
    - Determining Stream discharge will be used with GIS/Landuse Research
- Mobile GIS/GPS Applications
  - Compact field computer that integrates a mobile GIS, photo capture and a GPS receiver with 1 to 3 meter post-processed accuracy
  - Used to validate in the field accuracy of land use and hazard maps, to locate riparian buffer boundaries, and to create stream water quality maps
  - Real-time readings and map development with data integration
- Wesley College GIS Course
  - Computer simulation models (with and without a GIS interface) are used in instruction and research
  - New general studies GIS course developed and taught
    - Exercises for political science, nursing, homeland security, marketing, urban planning, ecology and business have been created
- Environmental Observation
  - Numerous additional environmental sensing instruments such as AUVs, Ocean Gliders, and the RV Sharp and monitoring the coastal and ocean environment.
- Environmental Monitoring
  - UD's Autonomous Underwater Vehicle (AUV) has been deployed at national and international off-shore sites
  - Remote control and real-time data analysis from Lewes, DE
  - *Art Trembanis and Matt Oliver (2010 NASTA PECASE Recipient)*
- NGS – Sustainability
  - Sequencing (and generation of other scientific data) is now outpacing Moore's law
  - Significant Costs to meet bioinformatics needs
    - Computational and database infrastructure
    - Trained Personnel
  - Cyber-based Collaboration has become one way of meeting the challenges
- *NECC Watershed Project: Miranda Lescaze, Vermont EPSCoR CWDD and Streams Project Director and NECC Watershed Project*
  - The Watershed Project: Engaging High School and Undergraduate Students in Water Research
    - 2011 NECC Annual Meeting March 15, 2011
  - What is the Watershed Project?

- Collaboration among NECC states' water outreach programs for workforce development and broadening participation
- Interaction, and exchange, of students from the NECC states, NY and Puerto Rico
- Approach
  - Use existing programs where possible to build a cohesive regional project similar to the VT EPSCoR Streams Project
  - NECC states can tailor this model to fit the strengths and needs of each state
- VT EPSCoR Streams Project
  - Collaborative effort by high schools, colleges and community partners around the state to collect water quality data on small streams.
  - Goals:
    - Workforce Development
      - Increase STEM
      - Recruit and retain students to STEM careers
      - Professional development for high school teachers
      - Recruit students to STEM majors
      - Increase the recruitments of female and under-represented minority students to STEM fields
    - Service to the state of Vermont
- Streams Project – Central Research Questions
  - How do land use patterns affect the physical condition and water quality of small streams?
  - How does water quality change in response to precipitation events? How might this response vary based on land-use patterns?
- Participants
  - High Schools
    - Collect biological, physical and chemical data in streams near their schools
  - Undergraduates
    - Work with high schools
    - Conduct laboratory analyses of macroinvertebrate and water samples or work on complementary research projects with community partners
    - Carry out independent research projects
- High Schools Program
  - Team structure
  - Vermont, New York, Puerto Rico & NECC states
  - 54 Streams high school teams since 2008 (48 VT, 3 NY, 3 Puerto Rico)
  - 2 Watershed Project high school teams this year
  - Training Week
    - Field laboratory and data analysis training
  - Monitor streams twice/month
  - Conduct a research project and present results at Spring Symposium
- Undergraduate Program
  - Interns from 12 Colleges and universities, to date

- 73 Streams Project interns since 2008 (52 VTm 21 Puerto Rico)
- 7 Watershed Project interns from NECC states this year
- Training Week
  - Field, laboratory and data analysis training
- Work in lab:
  - Water Quality Lab
  - GIS lab
  - Macroinvertebrate lab
  - Bacteria Ribotyping lab
- Pair with mentor to guide research
  - Four institutions
- Present independent research at Spring Symposium
- Database
  - The streams project data download website
- Current Monitoring Sites
  - Total sites: 165 (157 VT, 2 NY, 2 CT, & 4 PR)
  - Total streams : 66 (59 VT, 1 NY, 2 CT & 4 PR)
- Outcomes – Example Research Projects
  - Effect of Land Use on Benthic Macroinvertebrate Community Metrics in Vermont Streams (Saint Michael's College) – recently presented at ASLO annual aquatic sciences meeting in San Juan, PR
  - The relationship between Land Use and Phosphorus Loading: A sampling of Forested and Agricultural Land Use Basins Within The Lamoille Valley (Hazen Union School)
  - Analysis of phosphorus loading and microbial source tracking in the Lamoille River watershed, Vermont (Johnson State College)
  - Bioassessment of Wild Branch of the Lamoille River (Sterling College)
  - Stream Temperature Estimation Using GIS (UVM)
  - A GIS-Based analysis of the impacts of landscape level variables on water quality (Univ. of Puerto Rico)
  - A comparison of Hester-Dendy artificial samplers and brick samples of stream macro invertebrate communities (Saint Michael's College)
  - The Response of Stonefly and Mayfly Populations to Total Phosphorus Levels (Twin Valley Union High School)
  - The Effects of Rainwater on Stream Erosion and TSS Levels (North Country Union High School)
- Outcomes – Workforce Development
  - 100% of HS participants matriculated to college
  - 74% are women
  - 71% enter STEM majors
- What's next?
  - Watershed Project Spring Symposium April 26, 2011 in Burlington, VT
  - Streams Project 2011-12
    - 27 Internships

- 15 High school teams
- Watershed Project 2011-12
  - Involving NECC undergraduate interns and high school teams in training weeks
  - Determining how we can adapt Streams Project model to meet the unique needs of each NECC state

#### **E) NECC Committee Meetings (breakouts)**

#### **F) Lunch, networking, poster session**

#### **G) NECC State Projects: Fiber, Data Centers, Sustainability, and Leveraging:**

- *Rhode Island Fiber Progress and NECC Website:* Jennifer Specker, Assoc. PD, RI EPSCoR, and Zahir Shaikh, RI INBRE PI, and NECC Executive Committee members
  - Cyber-infrastructure in Rhode Island
    - A Collaborative Network Initiative
  - The OSHEAN State
    - Ocean State Higher Education Economic Development and Administrative Network
    - Founded in 1999 by URI, Brown, RINET
      - An established 501c3
      - Initial mission to increase access to internet
    - Serves High Education, Health Care, Government Agencies, Community
    - Began buying IRU's during 2003-2004
  - Network Design Criteria
    - Ultra-high speed networking to researchers while maximizing taxpayer value.
    - Resilient design
    - Must connect to OSHEAN
  - Funding a Collaborative Network Initiative
    - URI
      - EPSCoR – Dark fiber
      - INBRE – Dark fiber
      - DoE – Dark fiber & high speed wireless (802.11n)
    - OSHEAN
      - BTOP (funded by NTIA)
        - Was not successful during the first funding round
    - WCRPC
      - Washington County Regional Planning Council
      - Represents 9 Municipalities & 7 school systems
      - Ability to facilitate municipal gain/exemption space access
  - Ownership Models
    - Price various models
      - Do we build?

- If we build, can we build in the Muni space?
  - If so who will own it?
    - URI, OSHEAN, WCRPC
  - Do we buy an IRU?
    - Can we serve others in the IRU model?
- Two Steps Forward, One Step Back
  - Receives estimates from four vendors
    - Fibertech, Verizon, Sertex, Amtrack/Verosity
  - Ownership opportunities/issues
    - WCRPC
      - Each muni wrote letters of agency for URI/OSHEAN to use muni space
  - OSHEAN
    - OSHEAN's funded proposal was based on an IRU model – too risky to go request a revision
- Funding a Collaborative Network Initiative
  - EPSCoR ~ \$1.2M
  - INBRE ~ \$750K
  - DoE ~ \$924K (~\$500K wireless)
  - OSHEAN
    - BTOP ~ \$27.7M (+ \$10.7M match)
      - Existing fiber assets
  - WCRPC
    - BTOP \$9M BTOP Proposal (Underfunded)
- URI – IRU
  - URI decided an IRU approach was best
    - Issued an RFP for both build and/or IRU and received 3 Proposals
      - Cox – IRU only
      - Verizon – Build only
      - Fibertech – IRU only
  - Cox has tentatively awarded the contract
    - Targeting April 1st for a signed contract
  - Cox won OSHEAN's bid for RI fiber
  - IRUSs are 20 years with 4, 5 year extensions
  - URI and OSHEAN will swap fiber to extend each other's reach
  - OSHEAN's network ~ 340 miles
  - URI's network ~ 40 miles
  - OSHEAN to receive 48 strands of URI's 72
  - URI to receive:
    - 10 GB lambda connecting Kingston, Narragansett, Providence campuses
    - 1 pair fiber to Alton Jones campus
    - 2nd 10GB lambda researchers to access Brown's super computer cluster
    - Option to convert 2nd lambda to dark fiber
- Expected outcomes

- Researchers will upload their data within minutes instead of weeks
- Lab equipment will be more freely accessible remotely
- Real-time communication between researchers will be simplified and the quality improved
- Media created for/from research will be more readily accessible to other researchers and to future researchers.
- Increased ability to engage elementary and high school students with science.
- *Vermont Fiber progress:* Kelvin Chu, Assoc. PD, VT EPSCoR, NECC Executive Committee
  - Vermont Fiber Progress, 2011 NECC Annual Meeting
  - Vermont Bandwidth (2008)
    - IBM
    - State of Vermont
      - High Education
        - University of VT (200 Mb/s to Boston, 100 Mb/s to Albany, 45 Mb/s to Internet2)
        - Vermont State Colleges
        - 17 Private Colleges
  - NECC Overview
    - First plans launched among NE INBRE PIs in 2006
    - Face-to-face meeting to discuss fiber projects in Vermont in 2008
    - Collaborative EPSCoR Track-2, submitted Jan 2009 (\$6M)
    - Supplement requests to NIH-NCRR in Spring 2009 (\$8.4M)
    - NSF-C2 Awards to DE & RI in Sept 2010 (2.3M)
    - Total Awards to Date
      - NIH NCRR \$8.4M
      - NSF-EPSCoR \$8.3M
      - NECC annual meeting – 2010 in Burling, VT; 2011 in Orono, ME
  - Vermont Connectivity
    - 6 x 10 Gbps connections between UVM fiber ring and Albany
      - Position for USUCAN connectivity
    - 6 x 10 Gbps connections between UVM fiber ring and Boston (via NH)
  - 28 October 2010 Press Release
    - Over \$17M in competitive awards from the National Science Foundation and Nation Institutes of Health to the Northeast Cyberinfrastructure Consortium
  - Pilot awards for Novel Scientific Collaboration
    - 2010 Meiofaunal Metagenetics
      - Dr. W. Kelly Thomas Hubbard Professor of Genomics, Director of Genomic Studies; Dr. Holly Bik UNH, Dr. Jim Vincent UVM
      - Establish a database structure and data repository for eukaryotic metagenetic studies using the NECC data center
  - Skate Sequencing and Annotation
    - Cyber Enabled Research
    - Collaborative use of Specialized resources & expertise in an integrated process

- Skate clones MDIBL-Mount Deser Island Biological Lab (ME)
- Next Generation Sequencing: UD DNA Sequencing & Genotyping Center (DE)
- Sequence Analysis/Annotation: Analysis/annotation pipeline DE with ME, RI, NH
- Storage & Access of Sequence/Annotation data: UD shared data center (De)
- Storage & Access of Sequence/Annotation data: UD shared data center (DE)
- Public Dissemination: UD CBCB (DE) MDIBL (ME)
- Novel collaborative approach to next-generation sequencing
- NECC Vermont Summary
  - 1 Leg (BTV – ALB) in place;
  - 1 Leg (BTV – Hanover) coming in June;
  - Critical cyber-infrastructure across the Northeast Region to enhance national competitiveness of our researchers;
  - Novel scientific collaborations
  - Regional, cyber enabled collaborations
- *New Hampshire Fiber Progress: Scott Valcourt, Director, Project Management and Consulting Services, UNH and NECC Executive and Technical Committees*
  - New Hampshire CI Progress, NECC Annual Meeting, March 15, 2011
  - New Hampshire Focused Activity
    - Fiber
    - Leveraging
    - Sustainability
    - Data Centers
  - 'Big Green Panter' – NSF Segment
    - Fiber pathway runs from Manchester, to Plymouth, to Hanover, NH
    - Teljet Longhaul was selected at the fiber construction vendor and also as the IRU provider for the eventual 12 fiber strand for 20-40 years
    - Currently 15% of the ole application and make-ready work is complete.
    - Slated to be complete by June 2012
  - 'Southern Route' – NIH Segment
    - Fiber pathway runs from Manchester to Keene, NH, leveraging pathway to Hanover, NH.
    - Teljet Longhaul was selected as the fiber construction vendor and also as the IRU provider for the eventual 12 fiber strand for the next 20-40 years.
    - Currently 90% of the ole application and make-ready work is complete with 30% of the fiber segments already being hung on oles and strung in conduit.
    - Slated to be complete by June 2011
  - Leveraging BTOP Funding in NH
    - NSF & NIH segments are part of overall pathway associated with the BTOP project construction.
    - BTOP fiber will connect all USNH, CCSNH and all UNH Cooperative Extension sites to existing research sites in NSF & NIH
    - \$44.5 Million ARRA – funded project



- Data Collaboration and Integration
  - EOS-Webster provides a database of earth science data with complete search capabilities free of charge, and has a collection that currently sits at approximately 700,000 products.
  - VT STREAMS and NH have begun developing a storage and data visualization project that will integrate watershed data into EOS-WEBSTER
  - After the data has been validated and using Picture Post, data will be freely available for the community to perform analysis in the EOS-WEBSTER database and enhance knowledge discovery by researchers and citizen-scientists.
  - Dr. Annette Schloss, Lead
- *Delaware Progress and Data Center:* Karl Steiner, NECC Executive Committee, Sr. Assoc. Provost for Research Development
  - EPSCoR Cyberinfrastructure Workshop
    - First regional NENI (now NECC) Workshop held in Vermont in August 2007
    - Workshop held in Kentucky in October, 2007
    - Goal: Explore the State of Cyberinfrastructure in the EPSCoR States Process
      - Status reports by 24 EPSCoR Jurisdictions and four regional collaborations
    - Recommendations
      - Set aside of EPSCoR, OCI and/or co-funding budgets
      - Building on existing regional partnerships
      - Partnerships with other agencies such as the NIH/IDEA program
      - International partnerships, such as with the CANARIE network of Canada
  - NECC – Delaware Progress
  - NECC Cyberinfrastructure Upgrades in Delaware
    - Enhances Campus Cyber-Infrastructure at each of the Delaware EPSCoR Partners
      - UD: Bandwidth Expansion to Lewes Campus (100 Mbps to 1,00 Mbps)
      - UD: Regional Datacenter at DBI & Environmental Monitoring in Lewes
      - UD: Campus Backbone Upgrade & Distributed Data Center in Newark
      - DSU: Redundancy loop across Dover Campus
      - Wesley: Wireless access across Dover campus
      - DelTech: Campus-wide cyber infrastructure upgrades at Stanton and Georgetown Campuses
  - NECC Scientific Focal Areas
    - NIH-INBRE Scientific Focus;
      - Little Skate Genome
    - NSF—EPSCoR Scientific Focus
      - Metagenomes of Algal Blooms
    - Delaware Contribution:
      - Genome Sequencing of Little Skate, Distributed Data Center
      - Annotation Workshops
    - Track2-C2 Delaware Focus
      - Delaware Environmental Monitoring & Analysis Center

- Bioinformatics Capability
  - UD Center for Bioinformatics & Computational Biology
    - >30 affiliate faculty from five colleges
      - Arts & sciences, Health Sciences, Engineering, Agriculture & Natural Resources & Earth, Ocean & Environment
  - Jefferson Chair in Bioinformatics and Computational Biology – Cathy Wu
  - Four tenure-track faculty positions and start-up funding
  - BioIT Core at DBI
    - Computing infrastructure: high-performance computer cluster, database server cluster, 3-D visualization studio
    - Cyber-knowledgeable personnel (system, database, networking)
  - Protein Information Resource (PIR)
  - DSU Bioinformatics Initiative
    - Two tenure-track faculty positions under NSF-EPSCoR funding
      - Biostatistics faculty started in January 2010
      - Biostatistics faculty started Summer 2010
- Sequencing Center Upgrades
  - UD was awarded with NSF MRI grant to purchase a Pacific Biosciences RS Sequencer (Pacbio) – NSF plus UD cost share (\$1,062k)
    - Allows for high-throughput long-read length sequencing of single DNA molecules
    - The Pacbio will be effective for building a de novo genomic scaffold of the Little Skate genome, and resolving repeats and large genomic variations
    - Ideal complement to the Illumina sequencing platform.
  - UD has secured funds to upgrade the current Illumina GAIIx to an Illumina HiSeq2000 – purchased with NSF EPSCoR funds and UD cost share (>\$600k),
    - The HiSeq2000 will generate up to twice the number of reads at no extra cost per run.
- NGS—Data Analysis
  - Short Reads
  - Alignment/Mapping/de novo assembly
  - Analysis
  - Genome browser
  - De Novo Assembly (NECC)
    - Skate [CLC Bio]
    - C. elegans [CLC Bio, SOAPdenovo, Velvet]
    - S. invicta (Fireant) [CLC Bio, Soapdenovo]
  - Genome Structural Variants Detection (Larry Coghurn)
    - Chicken: FL-GS F0 & LL-GS-F0
    - SNP & Indel [CLC Bio]
    - CNV [CNV Seq]
  - RNA Seq (Carl Schmidt)
    - Chicken: lymphoma, control spleen, kidney, ileum, jejunum

- Exon identification and discovery, Gene level expression, Transcript level expression [CLC Bio]
  - Statistical analysis – identifying differential expression and feature clustering (to be done by [CLC bio])
- Delaware Environmental Monitoring and Analysis Center – DEMAC
  - Ingest Data from Diverse Sensing and Monitoring Platforms
  - Environmental Informatics Specialists design appropriate database structures for storage, analysis and dissemination of environmental data streams.
    - Diverse research projects will benefit from integrated environmental data streams including environmental modeling and process studies
    - Data will be available common format to researchers, educators and government agencies.
- Cyber-based Cloud Services
  - Amazon Web Services (AWS) provides cloud-based storage, web services and computation.
  - Skate project has utilized AWS-based interfaces for collaboration and teaching
    - Amazon S3 (simple storage service) provided a route for data exchange prior to Data center)
    - Amazon EC2 (elastic computation cloud) used for providing a web-based project BLAST interface.
- NECC Distributed Data Center
  - Data sharing capability to promote collaborations
  - 20 TB storage hosted at U. Delaware
  - Backup site at U. Maine
  - Web interface (U. Vermont)
  - Skate Sequencing Project initial user
- Increased Use of Bandwidth
  - Upgraded Connection from UD Campus in Newark to Lewes Campus from 100 MBps to 1,000 MBps in October 2010
- *Maine Fiber progress and Data Center:* NECC Technical Committee Co-Chairs Jeff Letourneau, Executive Director, NetworkMaine, University of Maine System, and Bruce Segee, Butler Associate Professor of Electrical and Computer Engineering, University of Maine
  - NECC State Projects: Maine Fiber, Data Cents, Sustainability and Leveraging
  - MaineREN Expansion
    - All Pole applications have been filed
    - Make Ready work in progress
      - Fiber construction driven by Utility companies completing make ready work
      - Ring close by the end of the Calendar year
      - Expansion into Northern Maine by May '12
      - Completion of all three Rings by Dec '12
    - Initial optical equipment installation in progress
      - Expanding existing node to handle new fiber routes

- Fiber Connected Entities
  - Bates College, Bowdoin College, Colby College, College of the Atlantic, The Jackson Laboratory, The Maine Public Broadcasting Network, Mount Desert Island Biological Laboratory, University of Maine, University of Maine at Augusta, University of Southern Maine, WC Community College, 12 High Schools.
- Data Center Sharing
  - Maine is operational as the backup data center
  - Technical personnel in Maine, Vermont, and Delaware
  - Reliable, redundant data
  - High availability of compute resources
- Other noteworthy activities and discussions
  - Federated identity management-recognize redentials from each other's institutions
    - Attended InCommon CAMP workshop in RI
  - Cloud computing-Model based on Amazon Cloud deployed in Maine has been successfully used in production
  - Business involvement- investments by state of Maine will bring more business and industrial traffic to Maine data center.
- *NECC Data Management: Ben King, Maine INBRE Bioinformatics Core, NEBC*
  - NECC Data Management: The NECC Data Sharing Policy
    - Benjamin King
    - Mount Desert Island Biological Laboratory
    - North East Bioinformatics Collaborative
  - NECC Data Sharing Policy
    - Policy for how data generated by the NECC will be:
      - Shared
    - Who can access what and when
      - Archived
    - Developed for Skate Genome Project, but serves as a guide for future collaborative projects within the NECC
    - Conforms to NSF and NIH Data Sharing Policies
    - Approved by NECC Executive Committee
  - Considerations
    - Comply with NSF & NIH Data Sharing Policies
    - Allow for validation prior to release
    - Follow what community has done in the past
    - Leverage existing public repositories for long-term data archiving
  - NECC Data Sharing Policy
    - NECC Researchers have access to all data
      - Encourage publications by NECC researchers
      - Includes collaborators outside of NECC (co-authorship)
    - Publications allowed prior to final paper
    - Proper acknowledgment of grants

- Exceptions are reviewed by NECC Executive Committee

#### **H) NECC Evaluation and Assessment:**

- *Special Report from the Executive Committee:* Kelvin Chu, Assoc. PD VT EPSCoR, NECC Executive Committee
  - Assessment and Evaluation, 2011 NECC Annual Meeting
  - Assessment
    - Requirements of NSF EPSCoR Track-2 and NIH NCRR Supplements;
    - ARRA funds
      - Quarterly reports
      - Annual Progress Reports
  - 5 Jurisdictions
    - Delaware: Karl Steiner
    - Maine: Vicki Nemeth (EPSCoR), Patricia Hand (INBRE)
    - New Hampshire: Scott Valcourt (EPSCoR)
    - Rhode Island: Jennifer Specker (EPSCoR), Zahir Sheik (INBRE)
    - Vermont: Judith Van Houten
  - Upcoming Evaluation
    - Maine:
      - INBRE supplement - EAC in August 2011, AAAS panel in fall 2011
      - EPSCoR Track-2 reviewed as part of Track-1 process;
    - NH: May 2011 Assessment of fiber transmission for southern route
    - RI: supplement AAAS panel assessment (2010), next review in 2012;
    - DE:
      - INBRE - next EAB April 2011, and INBRE AAAS review in Fall 2011
      - EPSCoR - RII AAAS Review in Fall 2011
    - VT :
      - NECC overview by AAAS panel in June 2011 as part of Vermont
  - Report Template
    - Major assignments only. All states have input to all parts
    - Project Summary
      - General summary by VT;
    - Intellectual Merit
      - NEBC – metagenome, Little Skate;
    - Broader Impacts
      - DE, ME – annotation jamborees;
      - DE – cybertools;
    - Programmatic Terms and Conditions
      - Broadening Participation (table)
      - Increasing Institutional Engagement
      - Describing leverage
    - Progress on Program Elements
      - Status (cyberinfrastructure landscape);

- Leveraged opportunities and activities (VT + state input);
- CI-enabled science and engineering projects;
  - VT – metagenome project
  - ME – Skate Genome Research
  - VT – RACER overview; NH – 1<sup>st</sup> RACER award;
  - RI – web-site hosting and RACER submission
  - ME – Data Sharing
- Diversity Plan
  - VT – Watershed project
  - DE – Diversity and Workforce development (next generation of bioinformaticians!)
- Dissemination Plan (VT, ME)
- Logic Model (VT)
- Sustainability Plan (Develop at March meeting)
- Management and Coordination Plan (VT)

#### **I) NECC External Engagement:**

- *Communication, Outreach, and Dissemination Partnerships:* Lillian Gamache, Project Administrator, VT EPSCoR, and Vicki Nemeth, Assoc. PD, ME EPSCoR Director, and NECC Executive Committee
  - 4.5 Dissemination and Communication Plan. A clearly articulated plan to communicate results, benefits, and processes of cyber-enabled research and research-based education to all citizens to build scientific literacy and strengthen the educational and research capacity throughout the consortium.
  - Communicating Results...
    - Awarded September 1, 2009 – August 31, 2012
      - Each member state generated local print and web announcements about award
  - Press Coverage Beyond NECC
    - NECC was 1 of 3 awards featured in this article
    - Drs. Karl Steiner and Judith Van Houten interviewed
  - Communicating Benefits of...to all citizens
    - Vermont hosted Press Conference
      - October 28, 2010 with U.S. Senator Patrick Leahy in Burlington, Vermont
      - Press release was sent out (written by the UVM Communications Office with VT EPSCoR Office) to all media outlets in VT
      - Coverage followed in the print and broadcast media such as the Associated Press, The Dartmouth (College) Paper, Burlington Free Press, MSNBC, local television affiliates for NBC, ABE and CBS, Fox
  - Press Conference October 28, 2011
    - Message – relevance to the state, university priorities
    - 5 State Consortium
    - Lighting the cyber “Black-Hole” of the Northeast

- Timing of press conference coincided with the completion of one leg of the fiber project in Vermont – something tangible for all citizens
- Further outreach to all citizens...
  - Press Conference Continued
    - Vermont Public Television also taped the announcement so it would be available for on-demand viewing on the web
    - The press conference is available on the web at [www.uvm.edu/EPSCoR](http://www.uvm.edu/EPSCoR)
    - All the media coverage is available on the Vermont NECC website
- Communicating...processes of cyber-enabled research and research-based education
  - Year 1 Deliverables
    - **Website:**
      - *RI-hosted NECC website*
      - The website was created and publicized throughout EPSCoR jurisdictions and at national meetings. The NECC website was made available to the public on March 12<sup>th</sup> 2010. It is hosted on the uri.edu network and can be accessed at: <http://necyberconsortium.org>
- RACER: Regional Awards for Cyberenabled Research (RACER)
  - RACER RFP is now posted for Year 2. Announcement sent out on the NECC distribution email and disseminated broadly on a state-wide email list (ERIS) in Vermont.
  - Awards are \$10,000
  - Information pertaining to the NECC RACER awards is posted on the NECC website
  - RACER pre-proposals are due on April 15, 2011
  - Full proposals due July 15, 2011
  - Five pre-proposals for Regional Awards for Cyber Enabled Research was received in Year 1; these form interjurisdictional collaborations for research projects that will utilize the cyberinfrastructure improvements funded by this RII-Track 2 award.
  - W. Kelly Thomas – UNH – Meiofaunal metagenetics was awarded in collaboration with Jim Vincent, UVM
- In addition to the data sharing and scientific collaborations that are ongoing...
  - Year 1 communication and dissemination efforts continued
    - Annual regional NECC meetings – allow for in person meetings and sharing of progress
    - The 1st Annual Regional Meeting was held on March 12, 2010 in Burlington, Vermont. All consortium members attended and presented progress in their respective areas. NSF EPSCoR Program Officer Jennifer Schopf attended.
    - Regular NECC videoconferences of Executive Steering Committee, - regular communication between partner states.
    - Technical Committee, Watershed leaders with liaisons, water working group, metagenomics group.

- NECC videoconferences are scheduled 2 times per month and all groups have been meeting on a regular basis.
- Annual Progress Report – way of communicating directly to NSF EPSCoR office
- The Annual Progress Reports were submitted by Year 1 by each NECC member state.
- Year 2 Communications and Dissemination Activities
  - Year 2 Annual Meeting – Underway! March 15, 2011 hosted by ME EPSCoR
  - AAAS Consultants Visit – scheduled on June 14, 2011
  - 3 podcasts, one 30-minute television show made in partnership with Vermont Public Television and Maine Public Broadcasting
    - Planning with VPT and MPB for episodes
    - Programming will air sometime in Year 2 and 3
- strengthening the educational and research capacity throughout the consortium...
  - What's next?
    - New Annual Report Template – APR due June 30, 2011
    - Continued Web Updates and expanding social networking visibility e.g. Facebook, YouTube, Twitter
    - 2nd RACER RFP Announced and Disseminated. Decisions available in late summer 2010.
    - Enhanced connectivity and collaboration will allow for increased data sharing and potential partnerships with other groups such as citizen science groups
    - Continued print (newsletters), web and broadcast media exposure

#### **J) NECC Committee reports and updates:**

- NEBC
- Technical Committee
- Water Research/RACER Committee
  - Regional Awards for Cyber-Enabled Research (RACER)
    - \$10,000.
    - RFP online
    - Required: Collaboration between at least 2 jurisdictions.
    - Encouraged: Water-related environmental research or bioinformatics, use of shared data facilities;
    - Pre-proposal process
    - Important Dates:
      - 1 March – RFP released
      - 15 April – Preproposals due (& available)
      - 15 July – Proposals due
    - Merit review by Water Research Committee
- Executive Committee
  - NECC Executive Committee Report



- Meetings of the Executive Committee
  - 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.
  - Took recommendations from the Water Research Committee on RACER awards and the selection of RACER awardees; approved year 2 RFP
  - January 6, our videoconference included Dr. Jennifer Schopf who discussed the annual report of progress for the NSF Track-2 grant.
- Presentations by Executive Committee Members
  - We had the opportunity to present our progress on the fiber network and cyber-enabled research at three venues in fall 2010:
    - October 4, 2010: Karl Steiner, 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, VA
    - 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 Labs
- The Executive Committee receives reports from the committees of NECC, such as the Committee that organizes and operates the Watershed Project for Workforce Development and Diversity
- Reviewing Progress over 1.5 years
  - Karl Steiner has given a flavor of where we were in 2006
  - What we were trying to accomplish:
    - Provide cyberinfrastructure for research and education in the NE region
    - Human infrastructure to make a virtual organization for distributed bioinformatics and data analysis
    - Pilot projects to develop the regional expertise to analyze genomics data, especially deep sequencing data, in a distributed manner; pilot projects to foster collaborations across the region
    - Collaborations on cybereducation, cyber-knowledgeable workforce development and diversity and outreach
    - Leverage resources to accomplish our goals
- How Far Have We Come??
  - Fiber projects are in various stages; some have been built
  - Carry out the pilot 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

- Pilots projects to inspire novel collaborations on water and cyberenabled research in the NECC region
- Collaborate on the Watershed Project to leverage the water education outreach programs in the Region for workforce development and diversity
- Foster novel collaborations that would not have happened without NECC and its cyber-infrastructure
- Examples of small businesses retained because of the new fiber and promise of fiber
- Novel Collaborations in Cyber-enables research that would not have evolved without the NECC
  - The Open Access shared database resource for eukaryotic metagenetics research (NH, VT)
  - Virtual Organization for bioinformatics and data analysis (NEBC, 5 states)
  - EOS Webster – GIS and positional information overlaid on field sample data for water research (NH and VT)
  - 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
  - Watershed Watch and Watershed projects collaborate in RI and NECC at large
- Topics for the Executive Committee Going Forward
  - We are no beginning to benefit from the Fiber Network: RAPID award to Kelly Thomas that required the new NH fiber connection to ME
  - Discussions of opportunities through grant proposals and applications to NIH, NSF and other agencies
  - Strategies for NECC and the next Track-2 round of proposals
  - Sustaining and growing NECC into the future

### **K) Planning for Year 3:**

- Next Track-2
- What can we do now that we could not do before?
- Moving forward with additional collaborative research projects