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 Winniepesaukee, 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, Date 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

o 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

o 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 dependent 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 Disseminaton: 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 becomes 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 loved into Excel and uploaded to Water Resources Agency's webpage for community view http://www.ipa.udel.edu/wra/
    - Podcast of Environmental Career Day, Sensors and Seminars available to NECC EPSCoR Community (www.denin.udel.edu/podcats)
    - Future Plans:
        - Currently 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/Land Use 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 macro-invertebrate 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 reams 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
        - Macroinverebrate 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 Macro-invertebrate 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/Veroxity
    - 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 OSEHAN will swap fiber to extend each other's reach
    - OSEHAN'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, Narraganseet, 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 Desert 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 Panther' – 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 Longhual 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 poles or 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) 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 demonic 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 Cogburn, UD)
        - Chicken: FL-GS F0 & LL-GS-F0
        - SNP & Indel [CLC Bio]
        - CNV [CNV Seq]
    - RNA Sequencing (Carl Schmidt, UD)
        - Chicken: lymphoma, control spleen, kidney, ileum, jejeunum
        - 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 100MBps to 1,000Mbps 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 credentials 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-1process;
        - 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 – 1st 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
        - 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 12th 2010. It is hosted on the uri.edu network and can be accessed at: http://necyberconsortium.org
- RACER: Regional Awards for Cyber-enabled Research (RACER)
    - RACER RFP is now posted for Year 2. Announcement sent out on the NECC distribution email and disseminated broadly on a statewide 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 inter-jurisdictional 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 for Year 1 were submitted 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 <u>distributed</u> 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 cyber-education, 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 cyber-enabled 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