UNDERGRADUATE GENOMICS RESEARCH THROUGH GCAT-SEEK: THE GENOME CONSORTIUM FOR ACTIVE TEACHING USING NEXT-GENERATION SEQUENCING

With Funding From

Vince Buonaccorsi, Juniata College

NSF/AAAS/HHMI: VISION and Change in undergraduate biology education

- Engage students in the process of science
- Present science as a "vibrant and active field"
- Authentic research experiences needed
- Stress core concepts and competencies

Genomics and V&C

Genomics and Bioinformatics approaches inherently offer high potential for forming student/scientist partnerships

But...
- High computational needs
- High sequencing costs
- High training requirements

Aims of the network

- Facilitate massively parallel sequencing research projects involving faculty teaching undergrads in their area of interest
- Coordinate group runs and/or negotiated group pricing with a number of sequencing centers and software providers
- Listserv connects liberal arts professors to experts in various fields
- MSI inclusion
- Increased access to shared computer resources for faculty and undergraduate students across the country
- Summer faculty development workshops which include funding for participant research projects
- Develop and publish genomics teaching resources

Affiliated Sequencing Facilities/Software Providers

Penn State
Ohio University
Indiana University
Delaware Biotechnology Institute
Cincinnati's Children's Genetic Variation and Gene Discovery Core Facility
The Network: 219 faculty from > 132 schools, 12 MSI
Most from small schools
Wide range of interests
Little next-gen practice
Experienced teachers

Juniata-HHMI COMPUTE CLUSTER
- One Master Node
  - Processors AMD Opteron Eight Core 16 Cores per Master Node
  - 16Gb RAM per Master Node
  - Hard Drive RAID Edition 3TB 6Gb/s Hard Drive, 7200RPM
  - Operating System CentOS 1
- Four Compute Nodes
  - Processors AMD Opteron 32 Cores per Compute Node
  - RAM 8GB/128Gb Ram per Compute Node
  - Hard Drive RAID Edition 500GB 6Gb/s Hard Drive

GCAT-SEEK Workshops 2013/14
The Staff
- Four breakout sessions:
  - Metagenomics: Gina Lamendella, Juniata College
  - Prokaryotic genomics: Jeff Newman, Lycoming College
  - Eukaryotic genomics: Vince Buonaccorsi, Juniata College
  - RNAseq:
    - Mark Peterson, Juniata College/Viterbo
    - Arthur Hunt, U KY
  - Pedagogy: Nancy Trun, Duquesne U
  - Assessment: Tammy Tobin, Susquehanna U
  - With six student TAs gives <5 to 1 facilitator/participant ratio within breakout sessions

Projects supported by network
- 2013 workshop: 13 projects for 20 faculty & 10 student participants
  - 7 RNAseq: 4 HiSeq Lanes
  - 2 Euk Genomic + 1 Prok Genomic: 1 MiSeq Run
  - 3 Metagenomic: 1 MiSeq Run
- 2014 workshop: 12 projects for 23 faculty and 8 student participants
  - 7 RNAseq: 6 HiSeq lanes
  - 1 Euk Genome, 1 RADseq, 1 Prok Genome: 2 HiSeq lanes
  - 2 Metagenomics: 1 MiSeq Run
- 2013/14: Non-workshop runs
  - 1 RADseq, 1 Euk Genome: 1 HiSeq Runs
  - 70 bact genomes: 3 Miseq Runs
  - 7 metagenome projcts: 5 MiSeq Runs

Workshop content
- Skills: Library prep, Expt design, Unix, Cluster computing, NextGENe, R

Bioinformatic Approaches: Euk Genomics
- Genome Assembly: Unix - Quake, SOAPdenovo2
- Genome Annotation: Unix - Maker
- Genome Comparison: Web - CoGe
- RADseq: Unix - Stacks, R studio
Other GCAT-SEEK facilitators

RNAseq: Mark Peterson
DGE: RSEM (Align to ref), DESeq, Networks or Functions
Variant Analysis: VarScan (ID variants)

Bacterial genomics: Jeff Newman
NextGENe, RAST

Metagenomics: Gina Lamendella
Usearch, Uclust, QIIME

Workshop Outcomes

- Workshop content and participant examples available at gcat-seeck.org
- Data available for network members (educational use only)
- Marked increase in
  - Knowledge and skills (MC test)
  - Enthusiasm for next generation sequencing
  - Confidence in ability to generate, analyze and incorporate NGS data into their research and coursework

Example: Juniata Bio III lab

- Question: Can we find candidate genes for basis of extreme longevity in rockfishes?
- Annotate a scaffold containing the same longevity gene from closely related long-lived and short-lived rockfishes using MAKER pipeline on Juniata Cluster
- Compare annotations and sequences using CoGe
  - Identity of putative aging genes helps narrow targets
  - Characterize divergence in coding and non-coding regions
- 3 x 4 hr labs ~ 100 students

Example: Gina Lamendella’s Metagenomics Research CURE

- Study Design, sample collection and water filtration
- DNA/RNA extraction
- 16S rRNA gene amplification
- Sequencing
- Multivariate statistics and correlation to metadata
- 16S rRNA gene analysis
- Alpha Diversity
- Beta Diversity

Outcomes

- Students learn a lot about what is happening in black box
- They become better troubleshooters and critical thinkers
- Engaged since they are working on applied projects
- Research publications
- Fellowships
- Job, grad school preparedness

Impacts of Network

- Student Involvement (using protocols, computers, data)
  - 2011/12 ~ 100
  - 2012/13 ~ 300
  - 2013/14 ~ 650
- Cluster: 6900 worker node jobs since summer 2013, over 200 users
- Publications
  - Web site (gcat-seeck.org) for breakout session manuals
  - CourseSource for published peer-reviewed manuals (coursesource.org)
  - 10 scientific, 2 educational publications submitted or accepted in three years from network
Next?

- Negotiating affordable HiSeq 2500 rapid runs
- More Course Source pubs
- Finish integration with iPlant Atmosphere
- Managing our initial datasets for broader use by network members
- 2 Workshops 2015…

Two Free Workshops 2015

- June 8-12 (Juniata) and 15-19 (Morgan State)
- Applications due Jan15
- Apply in teams
- One sequencing project supported per team
- For faculty teaching undergrads
- Sign-up for listserv under membership link at gcat-seek.org
- Request app or details from me (buonaccorsi@juniata.edu)