NGS: A Radical Transformation of Life Science

Charles D. Johnson, PhD
Director- Texas AgriLife Genomics and Bioinformatics
Managing Director – Center Bioinformatics & Genomic Systems Engineering
101 Gateway, Suit A, College Station, TX
979-862-3287
Charlie@ag.tamu.edu
www.txgen.tamu.edu
Key Points – Big Data

- Current State of DNA Technology
  - 1st Human Genome
    - 13 year/$3.8B ---> $1,000

Niall J Lennon @NJL_NGS · Nov 4
In the last 30 days we have produced sequence data for a 30X Human Genome every 32 minutes! @broadinstitute #HiSeqX

- Genomics explosion at Texas A&M
  - > 300 faculty working in genomics

- Challenges
Massive Acceleration in Reduction in Sequencing Cost

Cost per Raw Megabase of DNA Sequence

NGS

HGP

TAMU (GAIIx)

HiSeq (5x)

HiSeqv4

Moore's Law

NIH National Human Genome Research Institute
genome.gov/sequencingcosts

HiSeq 4000
Current State of Art

• 18,000 people per year per HiSeqX (10 pack) 30x coverage. (Only Human)
  – ~$100M to sequence and store data

• ~201 HiSeqX – 362,000 HG/year

• Every new born in US - $3B/yr
Sequencing Future
Why DNA Sequencing

~3 Billion Letters in HG (ATGC)
Length of your DNA - 4 round trips to the sun
Applications

• Basic Research
  • >98 “Seq” methods and counting
    https://liorpachter.wordpress.com/seq/

• Medicine
  • UK 100K human genomes
  • US 1M human genomes

• Agriculture
  • Marker Assisted Breeding

• Biology
A&M Highlights

- Human & Animal Health
- Plant/Animal pathogens
  - Host/Pathogen Relationships
- Plant & Animal Breeding
  - > different 17 species
- Over 3500 Bacterial genomes
- 1st Quarter horse sequenced
- 1st Scarlet Macaw sequenced
- 1st Pacific Shrimp transcriptome
- Insects, Companion Animals, Wildlife
- Bacterial Communities (Metagenomics)
Texas A&M Seed Grant Program

Empowering Genomics

$1.18M Seed Grant Program
Texas A&M Genomic Seed Grant

The Genomic Seed Grant is a two-year program with funds provided by the VPR ($250k), HSC ($100k), and AgriLife ($175k), totaling $525,000 to fund next generation sequencing (NGS) and bioinformatics analysis through the Genomics and Bioinformatics Service.

- **Number of grants**
  - 121 Submitted proposals ($2,712,140 requested)
  - 23 proposals recommended for funding ($524,829)

- **Participants**
  - 284 PIs and Co-PIs (342 total allowing duplications)
  - 230 TAMU faculty participants
  - 55 participants from outside TAMU (other system members, and scientists from around the country/world)
  - 132 planning meetings with A&M faculty (5 weeks).
2014 Report

- >100 NGS related grants submitted
- Total value of grants submitted (2014) >$42M
- Five year total funding associated with NGS >$50M

2014 Stats
- Number of projects 252
- Number of samples >7000
- Number of pipettes used: >1,200,000
- Number of sequencing runs 117
- Number of lanes run >900
Novel transcriptome assembly and improved annotation of the whiteleg shrimp (*Litopenaeus vannamei*), a dominant crustacean in global seafood mariculture

Big Data Challenges

• Basic quantitative research
  • Statistics, mathematics, machine learning

• Computational
  • Hardware
  • Software

• Bioinformatics (problem)
  • Over 1000 open positions listed in the last 4 months
    http://cbgse.tamu.edu/jobs-openings/
    • Domain Bioinformatics Knowledge
    • Background - Computer science (programming), statistics/mathematics, Biology/Genetics
    • Exposure to NGS data
  • Training – NOT FOR EXPERTS ONLY!
Center for Bioinformatics and Genomic Systems Engineering

101 Gateway, Suite A
College Station TX 77845

Bioinformatics, Computational and Systems Biology Research in Life Science and Agriculture.

Director - Aniruddha Datta
datta@ece.tamu.edu