PBIO/BIO 381 – Ecological Genomics

01/18/17

Definitions:

* Genetic source of adaptations
* Approaches and tools for Relationship geno/pheno/environment
* Environmental change and adaptations

Questions:

* Genes expressed at certain times
* metagenomes and metabolic potential
* independence from genetic code in epigenetics
* patterns of diversity
* constraints in genomes

Methods:

* De-novo
* Differential expression analysis: RNA-Seq
* Microbial community diversity, 16S metagenomics
* Reduced genome, a sample, RAD-Seq

Processes:

Interest: speciation and hybridization

Melissa’s and Steve’s story

01/23/2017

Info update 1

Rubric

* Outline : can give handouts
* 20 minutes
* learning/ engaging activity
* use of board effectively
* take home messages (star)
* examples from literature (beyond those given on Bb)
* Glossary is built here

Melisa Pespeni : Intro info update

Outline

1. Advances in Seq. Technologies
2. Range of Applications
3. General Library Prep. Workflow
4. Sequencing-by-synthesis (SBS)
5. Other Technologies
6. Learning Activity

1) Sequencing Technologies

Human Genome Project: 2001-2003 ABI – Sanger, took 15 years and sequenced 1 genome, $3B cost

Development of Hi Seq Tten in 2014 by **Ilumina**. This tool 1 day and 45 whole human genomes and cost $1000 each. Process: Microarray size slide: Ilumina flow cell with 8 lanes were capillaries are sampled by billions.

2) Applications: **90% global data**, mostly Ilumina seq.

WGS (whole genome), RNAseq, ChIPseq (Chromatin immunoprecipitation, protein bound transcription factors), targeted/capture sequence (use of probe= stretch of DNA that you synthesize, therefore you need to know about the target)

3) Workflow

Kind of library prep to make:

* Where is genetic variation? (Phenotypes)
* # of samples: population, individual, comparative studies
* Model or not
* Demographic history of population
* Adaptive genetic variation
* Gene expression variation
* Determined by:
  + length of reads: longer reads will be easier to assemble
  + # of reads
  + distribution of the reads: knowing relative position helps with assembly, are they randomly distributed or not

Steps:

* Extraction: DNA or RNA -> **cDNA**
* Fragment sample
* Ligate adaptors: individual barcode
* Add sequence adaptors
* PCR amplification

Reduced representation

RNA: coding regions

GBS/RADseq: near restriction sites

4) SBS

* Flow cell , has lanes and oligos (eg. P7, P5) that attaches to your DNA that has adaptors that sample by ligation. Your DNA is attached to a sequence adaptor and may have an ID barcode. Then **bridge amplification** where the pieces bend over and amplify (to copy and make signal stronger). Bridge released and then cluster generation (clusters of the same sequence). For each cluster a snapshot is taken of the nucleotides, as the polymerase is adding the nucleotides.
* PacBio: Single Molecule in Real Time. Gives really long reads but accuracy is lower.

\* Drawings on notebook\*

6) Learning Activity

Glossary:

Ilumina reads: short (50bp), long (100,150,300 bp),

SMRT: extra long (10000 – 60000 bp). Can also be single vs. paired end.

----

Paper discussion: Ellegren, 2013

Omics: when is it useful?

* Can be used as an unbiased view

Storage for genomic data! Since it should be publicly accessible

Quality control: files are raw, quality processing is done afterwards

Table 1: 2,200 update of reference genomes of eukariotes

Great proportion of non-model organisms. Does not include transcriptomes.

How do we determine what is a reference genome? In scaffolds, not in chromosomes. There is a database to query. Sample size is 1 (not always) but how to know if it is a reference that works for a specific question.

Next Monday: 4 Info-updates