Ecological Genomics Notes

P. Alexander Burnham

January 23, 2017

January 23, 2017:

Info Update:

- Advantages in Seq tech.
- Ranage of applications:
  - WGS (whole genome sequencing)
  - RNAseq (seqencing RNA conveted to cDNA)
  - Targeted caputure seq. (string of probes mixed with sample, pulls immune related genes from organisim and seq just those)
  - Chipseq (chromatin imunoprecipitation seq, recognizezes and antibody and puls out all DNA bound to that portein)

Why one or the other?

Genetic variation

```
- phenotypes
```

number of samples

```
- population
- individual
- comparitive studies
- model or note
```

Demographic history

Adaptive geneitc varaitin

gene expression car.

length of reads

number of reads

distribution

Reads:

```
- short = 50bp
- long 100 bp, 150 bp, 300 bp (miseq)
- 10,000-60,000bp = SMRT
```

Single vs. paired end

- General library Prep. Workflow
  - extraction (DNA, RNA -> to cDNA)
  - fragment sample
  - ligate adaptors (indvidual barcodes)
  - add seq. adaptors

Reduced Rep

  - RNA -> coding
  - GBS/RAD-seq
  - near restriciton sites

- Sequencing-by-synthesis (SBS)

  - bridge amp
  - cluster gen.
  - labeled dNTP (ATCG)

- Other Technologies
- Learning Activity

Human Genome Project (2001-2003)

```
- ABI = Sanger
- 15 years
- 1 geneome (one person)
- $3 billion
```

Uses PCR and sequences broken by faulty base pairs to work backwards

2014 X-Ten releases

```
- Hiseq by Illumina (look up video of how it works)
- 1 day
- 45 whole genomes
- $1000 bucks each
```

Sheet of glass with 8 lanes with flow cells... (look up)

Take home messages

- Likely using Illumina seq. (just library tech. changes) usually SBS
- adaptors are markers (barcodes) used to identify samples during sequencing

  - first thing that's sample and gives ID barcode = seq adaptor
  - alago is a sequence of DNA that is attatched to plate binds to sample

- Model vs. non Model:

  - short reads (assembly to create long sequnce based on short reads that shift)
  - denavo assembly -> computer program (added variability with mixed sample)
  - 15% error for SMRT can be reduced to less than 1 with repeated passes
  - illumina is much smaller but more accurate (0.05% error)
  - combine the two to have a higher degree of confidnece

Paper Discussion:

Three advances in biology:

- Modern Synthesis (evolution (Darwin) and population genetics (Mendel))
- Watson and Crick (molecular biology DNA)
- Omics Era (genomics, preteomics etc.)

What do we think about this?

- Phylogenies -> reducing error bars or reshaping question?
- Do you throw out experimetnal design and scientific method for large scale shot gun blast sequencing? - > can still be used to do hypothesis driven science, but also something to be said for sending a teleoscope into deep space.
- Most journals and funding sources require all seq data be made public - leeds to a storage space issure -> genbank ran out of space and needed emergency funding from congress.
- data are not reviewed very well some gen bank sequences are kinda rough
- reference genomes are a sample size of 1 so it might not be representative

On Wednesday two papers to talk about:

1) where do genomics data have limitations and why we use them
2) discussion leaders (sign up via blackboard)
3) next week talk about some of library preps (four of them) four differnt update people

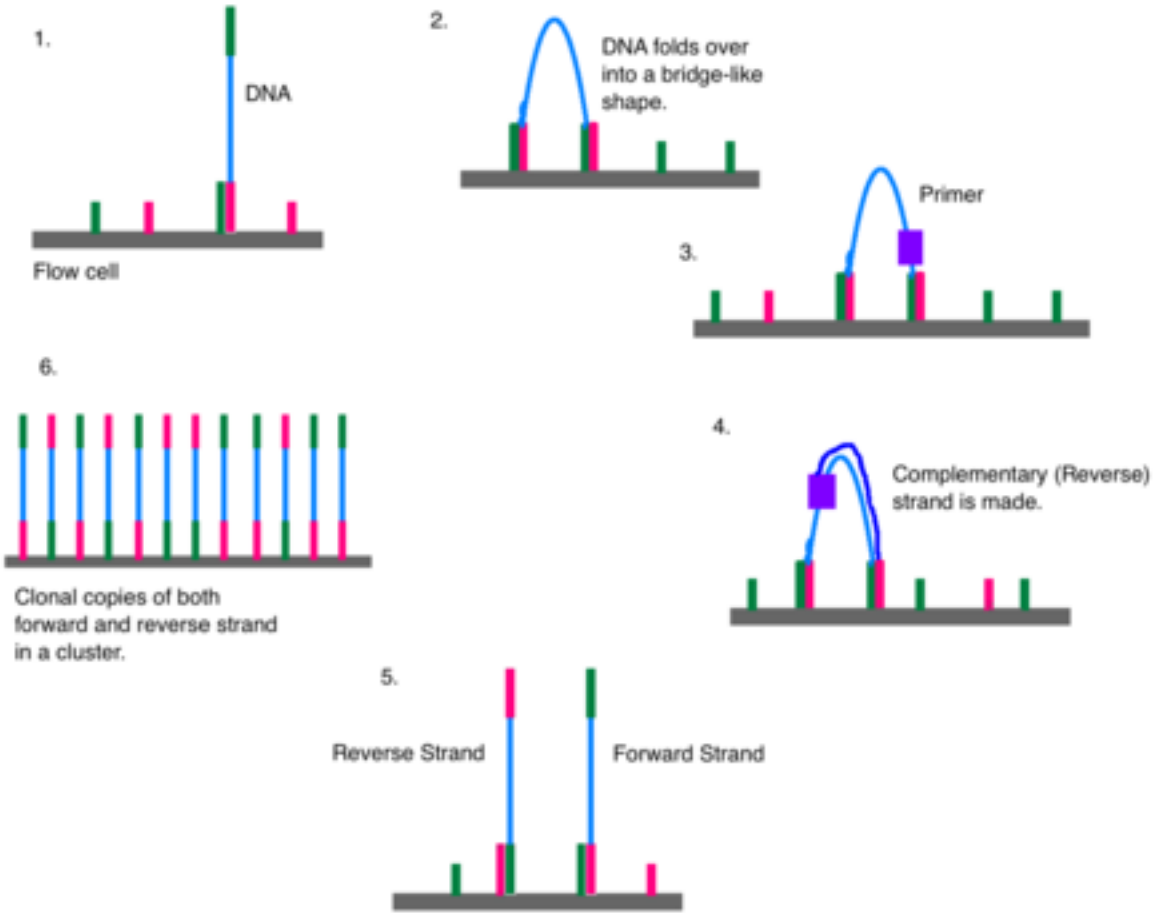

Figure 1: Illumina Sequencing