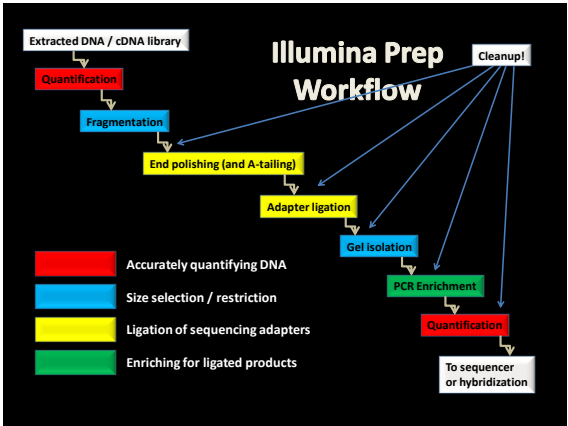


# Library Prep & Multiplexing



### Quantification Methods:

- Gel electrophoresis
- Nanodrop - spectrophotometry
- BioAnalyzer

### Qubit - fluorometry

**Benefits**

- pretty cheap
  - ~\$1000 initial investment
  - ~\$0.55 / sample
- fairly quick
- (usually) low sample consumption (1-2 ul)
- minimal contamination issues
- highly accurate readings

**Drawbacks**

- a little slower, more consumables, than Nanodrop
- should have ballpark [DNA] before running

### Fragmentation Methods:

#### Sonication:

**Pros:**

- Can be high-throughput (8-24 samples)
- Very little sample loss
- \$0 / sample

**Cons:**

- Can be slow
- High initial cost (\$19-40,000)

#### Nebulization:

**Pros:**

- Low initial investment: \$20 / chamber

**Cons:**

- Typically lose ~1/2 sample
- Can be slow (6 min / sample)
- Costs add up \$20 / sample

### Cleaning Methods:

#### Qiagen column-based cleanups

**Benefits**

- fairly rapid cleanup per sample
- affordable (~\$2 / cleanup)
- ~ required for gel isolation
- range of cleanup products
  - PCR cleanup kit
  - Gel Extraction kit

**Drawbacks**

- reliably mediocre
  - yield often on low side
- columns not amenable to high-throughput
  - high-throughput option available, but expensive

### Cleaning Methods:

#### AMPure paramagnetic beads (SPRI-beads)

**Benefits**

- fairly rapid cleanup per sample at high sample numbers
  - very amenable to high-throughput
- affordable (\$0.20-0.60 / cleanup)
- high yield
- preferentially binds fragments >200 bp

**Drawbacks**

- slower than Qiagen for low sample numbers
- technically a little more challenging
- initial cost (magnet) a little high (\$250 -> \$500)
- short shelf life (6 months to 1 year)
- Less efficient with very low [DNA]

**Size Selection:**

Gel Isolation – required step(?)

Procedure

- ≥2% agarose gel
- 1 hour run time
- UV illumination
- cut out gel pieces of selected size
- Qiagen gel cleanup

Tips?

- use shortened pipette tips to cut gel plugs out
- melt gel directly out of tips with buffer QG

**Size Selection:**

Alternatives?

A couple of possibilities

- alter AMPure buffer salt concentration through dilution
  - allows selection of variable size classes
- SKIP IT!
  - can be ok for single end preps

**Multiplexing**

Addition of a unique sequence identifier (barcode) allowing multiple samples to be run together on a single flow cell lane.

Multiplex level =  $\frac{(\text{Lane Capacity}) \cdot (\text{Correction}) \cdot (\% \text{ hitting target})}{(\text{Depth Desired}) \cdot (\text{Target Genome Size})}$

Straub et al 2012

**Multiplexing**

Extracted DNA / cDNA library

Quantification

Fragmentation

End polishing (and A-tailing)

Adapter ligation

Quantify and pool after adapter ligation

HOWEVER:

Maintenance of about equal [DNA] is strongly recommended.

Gele isolation

PCR Enrichment

Quantification

To sequencer or hybridization

Possible to pool before gel size selection:

Quantify → Pool Subset → Gel → PCR → Quantify → Final pool

**Hyb-Seq**

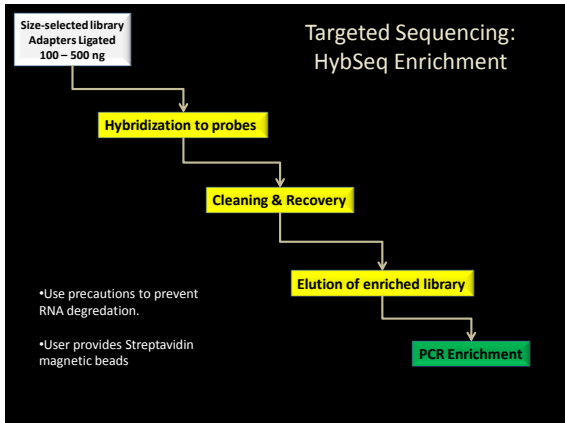
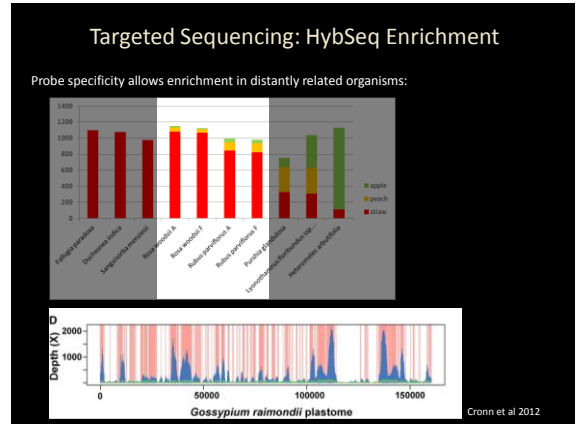
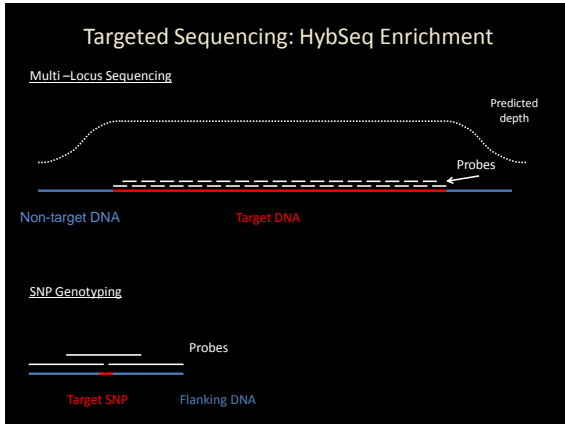
**Targeted Sequencing: HybSeq Enrichment**

**Solution Phase Hybridization**

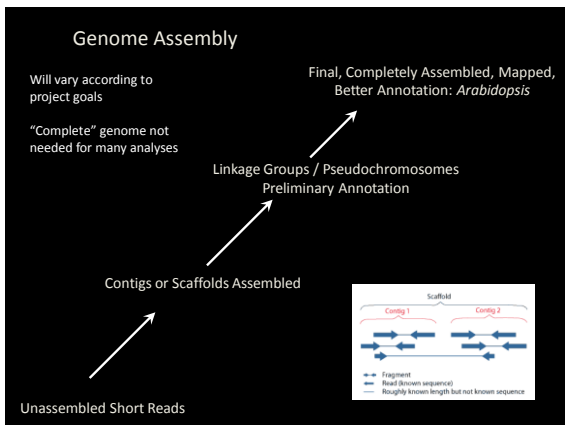
Examples: Agilent SureSelect, NimbleGen SeqCap EZ, Microarray MyBait, etc

- ‘Baits’ synthesized on arrays
- 120 bp RNA probes.
- Hybridization in solution
- Immobilization via biotin-streptavidin capture
- 1 – 5 ug DNA of input library.
- Scalable to large samples, targets

Slide courtesy of Rich Grimm



## Genome Assembly



- ### Assembly metrics: What is a good assembly?
- **Coverage:** Portion of the target genome represented in the assembly.
  - **N50:** Contig size where the sum of all equal and greater contigs equals 50% the sum of all contigs.
  - **Number of contigs or scaffolds:** A lower number indicates better success. \*
  - **Depth:** Number of reads hitting a particular position (often averaged across a region).

## Data Preparation

- Quality Control
  - Trim low quality bases at the ends of reads
  - Removal of adapters
- Removal of high-copy regions
  - Chloroplast
  - Mitochondria
  - rDNA

## Kmer selection

AGCTCTAGGA  
| | | | | | | |

- Kmer – A substring, with length K, of a read.
  - A read of length L has  $(L - K + 1)$  Kmers.
- Optimal value of K needs to be found for each assembly.
- K too low: Increase of spurious Kmer overlap.
- K too high: Kmers will not overlap with enough neighbors.
- Depends on sequencing depth, error rate, genome complexity

## Programs

- Velvet
- ABySS
- SOAPdenovo
- ALLPATHS-LG
- Major considerations (beyond quality)
  - Memory Use
  - Time
- Comparisons:
  - Assemblathon, GAGE

