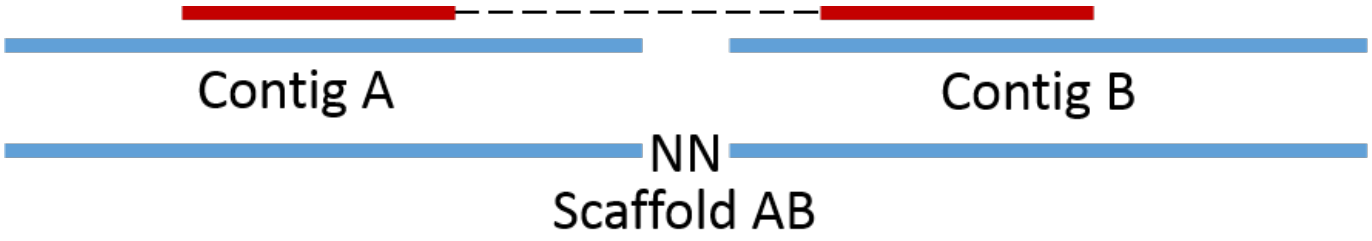
Mate-Pair Update

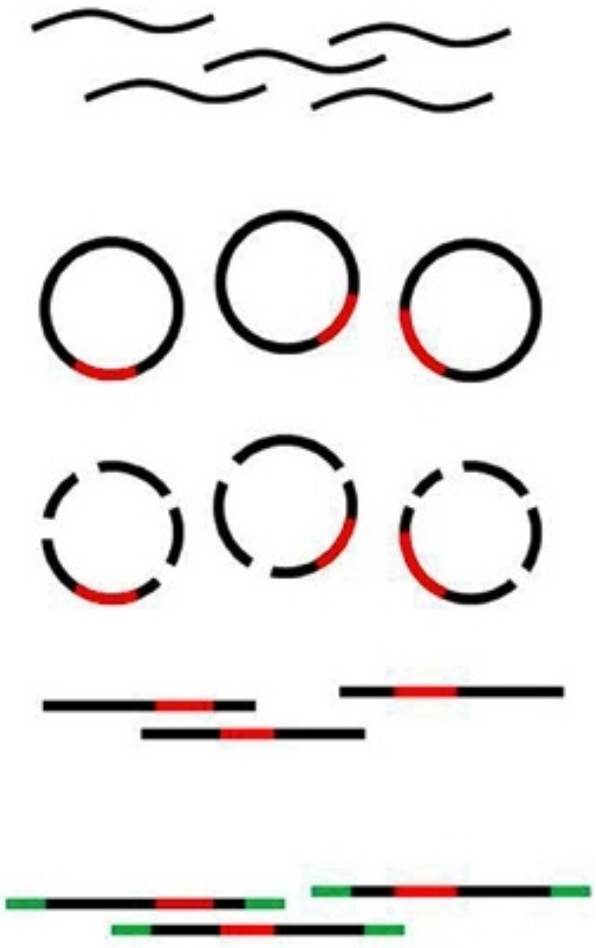
Brendan O’Connell
Steven Weber
Green Lab

Why Mate-Pair?

- Long distance information is required in order to scaffold together contigs generated by shotgun data
- The goal of making a mate-pair library is to generate reads that contain information about the ends of very long inserts (2-10kb)
- It is very important to size select this DNA to a tight size range
  - This gives us concise information about how many “N’s” exist between contigs



Molecular Biology Review



- DNA sheared and size selected to a specific range
- Biotinylated linker molecule ligated to both ends of insert, resulting in circularization
- Exonuclease of any non-circular fragments
- Shearing of circular DNA followed by Illumina library prep

Problems we have encountered

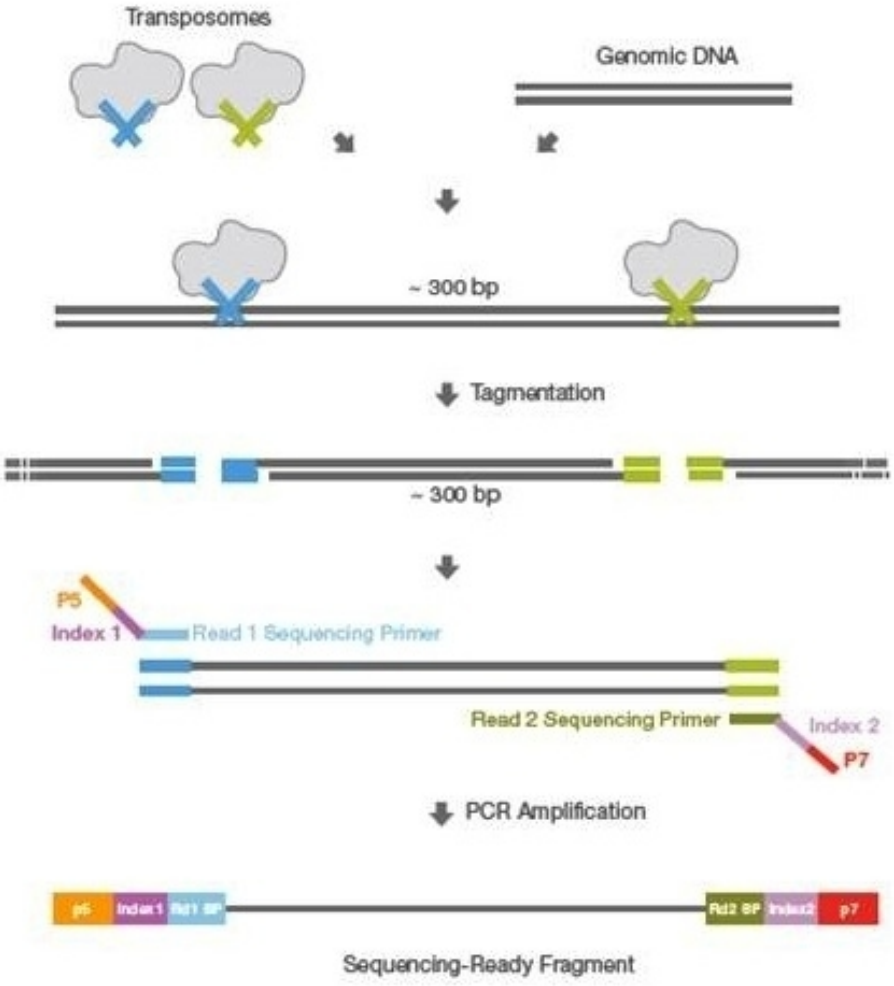
- Lucigen Mate-pair kit gave us very low complexity libraries
- One big problem is that a very small percent (5-10%) of your sheared DNA makes it through the tight size selection process
- Only 5-10% of this size selected DNA ever actually gets circularized with the molecular biology of the Lucigen kit
  - This means that only ~2.5ng out of every 1ug of DNA is ever actually circularized
- If your initial insert length is 4kb and your finished illumina library is 400bp, this means only 10% of your circular DNA actually contains junctions
  - This means that less than 1ng out of every 1ug processed are actual mates

Troubleshooting

- One way to compensate for this extreme loss is to start with a very large amount of DNA
  - This works as long as you have a very large source of DNA, which is not always the case
- Another way is to change the molecular biology to be more efficient
- We were introduced to a method of library prep using an enzyme called Tn5, or Tagmentase

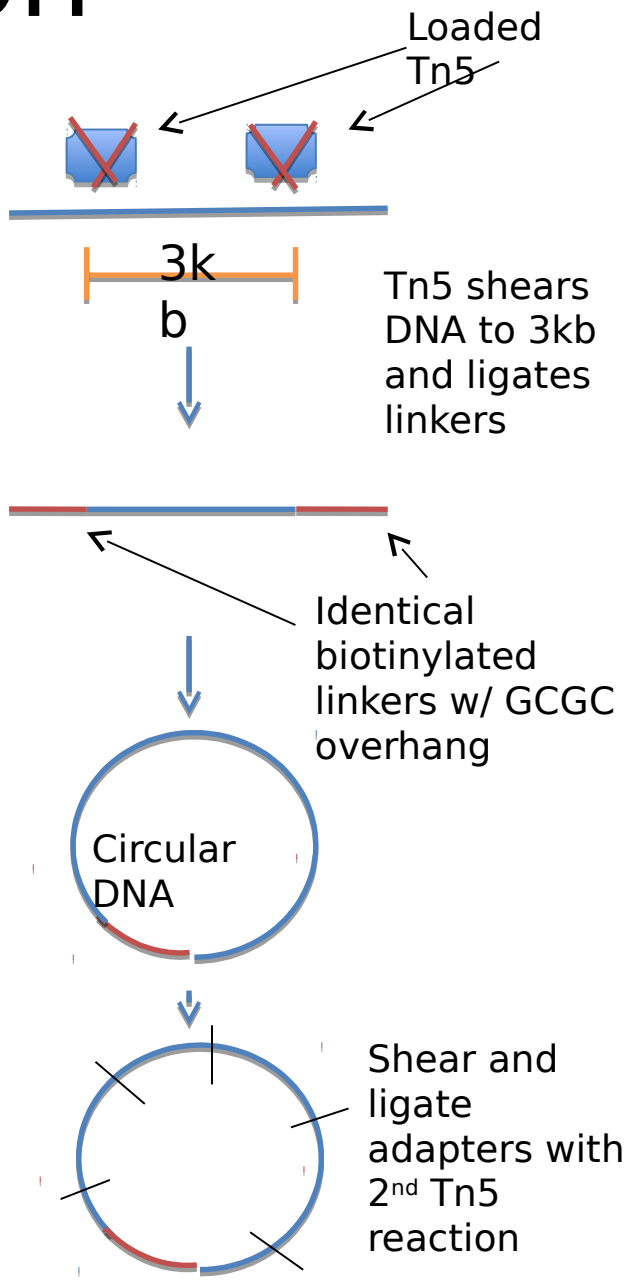
Tagmentase

- Tn5 is an enzyme that recognizes and binds a specific 19bp sequence
- In order to become an active “Transposome” it must bind two of these sites
- Exposing the Transposome to long strands of DNA causes it to cut and ligate the bound fragments to either side of the cut
  - In this case, we have Illumina adapters downstream of this 19bp recognition site being loaded into the Transposome
- This process replaces the shearing, end repair, adapter ligation, and adapter fill-in with just one reaction



Tagmentation

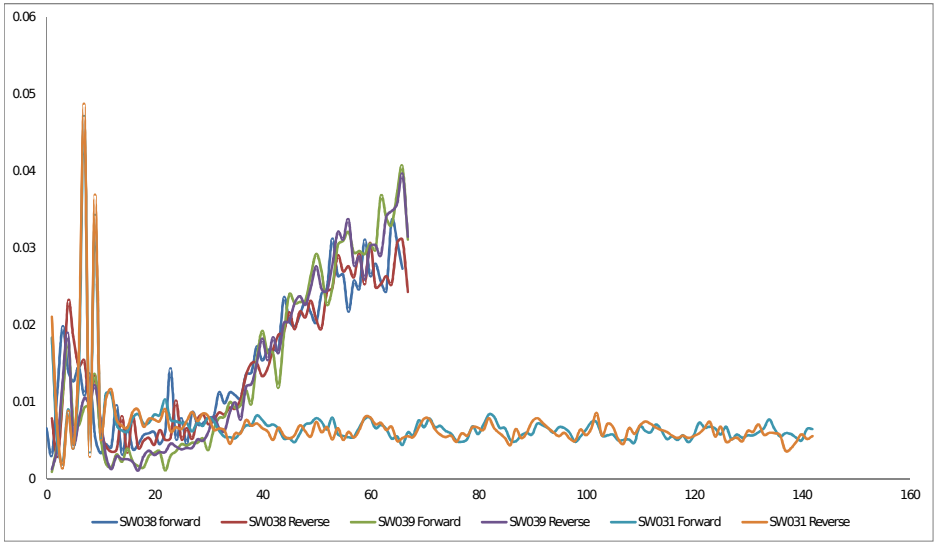
- Tn5 will load virtually any fragment that contains the 19bp recognition site
- If we design oligos that will be loaded by Tn5, contain biotin and a palendromic overhang, we can use it to shear our DNA to any size and then proceed directly to circularization
- Digest all linear DNA away and perform a second Tn5 reaction, this time with Illumina adapters loaded into the Transposome
- Biotin enrichment and PCR



Pros and Cons

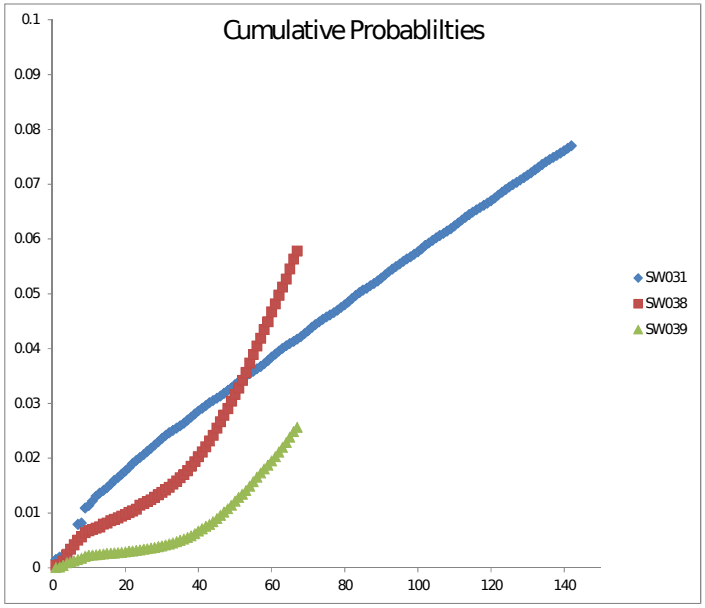
PROS

- Very fast and efficient, skips many clean-up steps
- Linker sequence found more commonly in the middle of the sequence



CONS

- Increased chimeric rate
- Longer linker sequence
- Not as fine-tuned as the shearing protocol



Analysis Discussion

- Throwing out reads that are not “true mates”
- Differences in linker sequence
- Saving reads where no linker sequence is found