**1) You mentioned that after digestion with restriction enzyme, fragments that are too small or too large would be left out of sequencing. Can you specify thresholds for this?**

The sizes of the fragments after RE digestion are ~ 250 bp – 1000 bp. We do not believe that they were subjected to any further size selection prior to sequencing.

**2) I want to confirm our understanding that even though the sequencing is paired-end, the data does not include approximate distance between the two reads of a pair. This is very valuable information generally available for WGS assembly, so we want to make sure**.

At this point, we have both real and simulated WGS datasets.

The real WGS shotgun dataset were sheared using two different methods: Physical and Enzymatic shearing. By aligning the physically sheared fragments to the reference genome (i.e., the BACs) we observe that the mean fragment size was 343 bp, with a standard deviation of about 110 bp. For the enzymatically sheared fragments, ~ 20% of the reads were chimeric so these were not used. so I didn’t use this dataset. In the future we will use physical shearing only. What size fragments do you suggest we use? And how much variation in size will be tolerable (too narrow and we won’t have enough DNA for sequencing).

In simulated WGS data, there are two different fragment size 750bp and 1,500bp. Hung-ying used this information and these data in SSPACE part.

**3) Is it possible the sequencer makes an error in reading the barcode, and thereby we assign it to a wrong BAC?**

The barcodes are two edit distances apart. The Q30 error rate of a Mi-seq 2x300 bp run is about 90%, which means 90% of the reads have < 0.1% error rate. Hence, although it would be POSSIBLE to mis-assign a read to the wrong BAC this would be rare.

**4) Are you using restriction enzymes with blunt cuts, or sticky? Does it matter? What percentage of chimeric sequences we should be prepared for?**

Sticky overhands. I don’t think it should matter. We are testing methods to decrease the % chimeric reads. We’ll let you know what we can achieve. For the current pipeline (where we computationally cut sequences that contain a relevant restriction enzyme) the chimerism issue isn’t a problem.

**5) You mentioned it would be ideal for you to move towards a single enzyme and partial digest. You also mentioned that you will cut reads when you encounter the restriction enzyme patter in the middle of a read, deducing it must be a chimera. With a partial digest, the latter scheme would not work.**

As mentioned above CURRENTLY because we are concerned about chimeras, we computationally cut any detected restriction enzyme recognition sites in the sequenced reads. On the other hand, we are exploring ways to substantially reduce the frequency of chimeric molecules.  If we are successful at this we will attempt to implement the partial digest strategy.

Yes, I would like to discuss with you guys. Do we setup a meeting at the same time this week?

Here is Hung-Ying’s skype ID "bigbear840904"