Assume there is a plant mitochondrial genome, whose size is 500,000 bp and whose genome is circular. We need to conduct de novo genome sequencing to get its genome sequence.

1. How can you get mitochondrial DNA? What problems do we need to concern for this step?
2. Paired-end or single end sequencing?
3. For DNA fragmenting, what sizes of DNA fragments will you use (or insert size) ? what kinds of libraries will you use?

A. 800bp

B. 1Kbp

C. 5kbp

C. use all above

1. What average coverage do you plant to get for sequencing? How many reads do you need to get for this coverage? How many lanes do you need if you use Illumina Hi-Seq 2000? Hi-Seq 2000 has 200-400 million paired-reads (100bp) per lane.
2. Which assembler will you use? Why?
3. What computer do you use for assembly?

A. 4GB laptop

B. 20GB workstation

C. computer cluster in HCC

1. According to your estimate, how long does it take for assembling?

A. 30 minutes

B. 1 hours

C. 10 hours

D. 4 days

1. What software do you use for scaffolding? How long does it take?
2. What is longest gap in one scaffold? How do you fill gaps?
3. How do you determine if your assembled genome is good enough?
4. How do you annotate genes?