Small genome *de novo* assembly

A plant mitochondrial genome,
whose size is 500,000 bp

How can you get mitochondrial DNA? What problems do we need to concern for this step?

- Two options:

(1) get mitochondrial DNA only (grow your plant in a dark room)

(2) get total DNA

Paired-end or single end sequencing?

- Paired-end sequencing is necessary for scaffolding and good quality.

- For DNA fragmenting, what size of DNA fragments will you use? (or insert size of paired-end sequencing)

  A. 800bp

  B. 1Kbp

  C. 5Kbp

  D. use all above

Usually, 1Kbp is enough.

coverage

- What coverage do you sequence? 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.

Coverage = 100x

Coverage = NL/G

N=coverage*G/L=100x500Kbp/100bp=500K

5% of total DNA is mitochondrial DNA, then we need 20 x 500K reads

Which assembler will you use?

- Velvet for small genomes

- What computer do you used to do assembly?

  A. 4GB laptop

  B. 20GB workstation

  C. computer cluster in HCC

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

  A. 30 minutes

  B. 1 hours

  C. 10 hours

  D. 4 days

What software do you used for scaffolding? how long does it take?

- Velvet or SSPACE
- Couple hours are enough

What is longest gap in one scaffold? How do you fill gaps?

- If we use 1kbp insert-size library for paired-end sequencing, the longest gap is around 1kbp.
- PCR is enough

How do you determine if your assembled genome is good enough?

- See if it is circular. If we do not get a circular genome, we may consider to have an additional library with a different insert size.
- Compare with known mitochondria genome

How do you annotate genes?

- Using existing software tools, such as Dogma, ORF finder, or T-RNA scan.
- Collect all known mitochondria genes, and use BLAST to search the assembled mitochondria genome against these known genes.