

Summary

- Sanger sequencing reads DNA via synthesis; 800-1000bp.
- Assembly Paradigms:
 - Shortest Common Superstring (NP-hard; sensitive to repeats)
 - Hamiltonian cycle in overlap graph (NP-hard)
 - Eulerian cycle in de Bruijn graph (polynomial in basic form, but large # of solutions)
- Overlap alignment can be computed with slight variant of sequence alignment DP.
 - K-mer hashing technique avoids all pairs overlap alignment