

90-Minute Discussion Session Plan

Lecture 4: Long-Read Assembly

Course: BINF301 – Computational Biology

Instructor: Tom Michoel

Date: 26/01/2026

Created with Copilot

0–10 min – Warm-Up

Prompt:

- “Which long-read technology (PacBio CLR, PacBio HiFi, Nanopore) do you find most intriguing, and why?”
- “What is one challenge you associate with long-read sequencing?”

Instructor note. Goal: Activate prior knowledge and normalize uncertainty around long-read technologies. Encourage students to distinguish *read length vs. accuracy*, and recall that early long-read technologies were noisy, while HiFi reads dramatically improved accuracy.

10–25 min – Think-Pair-Share

Main prompt: Why are De Bruijn graphs difficult to use effectively for long-read assembly?

Follow-ups:

- “What assumption in DBG construction breaks for long reads?”
- “Why is overlap-based assembly more appropriate?”
- “How do read count and error rate affect k -mer completeness?”

Instructor note. Students should recall that DBG assembly requires high and relatively uniform k -mer coverage. Long-read datasets have *fewer reads* for the same coverage, and historically had *higher error rates*, producing many low-frequency erroneous k -mers. Therefore long-read assemblers rely on overlap graphs, not DBGs.

25–45 min – Structured Group Discussion

Students form groups of three with the following rotating roles:

Roles

- **Summarizer:** Explain how PacBio HiFi differs from older PacBio CLR/Nanopore reads and why this affects assembly.
- **Questioner:** Prepare one question about error correction strategies (hybrid, hierarchical, direct).
- **Connector:** Link overlap detection techniques (minimizers, MHAP) to the computational difficulty of all-vs-all overlaps.

Starter Question

How do modern assemblers (Canu, Flye, HiCanu, HiFiAsm) differ in how they treat noise, repeats, and haplotypes?

Instructor note. Expected points:

- **Canu**: hierarchical correction; MHAP for fast approximate overlaps.
- **Flye**: repeat graph model; distinguishes bridged vs. unbridged repeats.
- **HiCanu**: optimized for HiFi; homopolymer compression; pileup-based correction.
- **HiFiAsm**: haplotype-aware assembly; multi-pass correction and string-graph phasing.

Encourage students to compare *overlap graphs*, *repeat graphs*, and *string graphs*.

Break (15 min)

45–65 min – Applied Exercise Block

See handout.

Exercise 1 – MinHash Overlap Detection

Sketch how minimizers or MHAP signatures help detect approximate overlaps between long reads.

Exercise 2 – Bridged vs. Unbridged Repeats

Given a repeat diagram, identify which repeats are bridged by long reads and which require Flye’s repeat-graph method.

Exercise 3 – Haplotype Bubbles

Given a string-graph bubble, determine whether it represents heterozygosity or sequencing noise.

Exercise 4 – Hi-C Scaffolding Clues

Interpret a schematic Hi-C heatmap showing either a clean diagonal or a diagonal break.

Instructor note.

Instructor should prompt students to think about:

- Evidence that a repeat is bridged (a single read spans both copies).
- Which SNP-rich reads inform haplotype assignment.
- How SALSA2 or YaHS break contigs at low Hi-C support boundaries.

Applied work reinforces understanding of algorithms and data interpretation.

65–80 min – Synthesis

Discussion prompts:

- “What makes long-read assembly fundamentally different from short-read assembly?”
- “Which problem does each major tool (Canu, Flye, HiCanu, HiFiAsm) solve particularly well?”
- “If you designed a new assembler, which component would you innovate—overlap detection, repeat resolution, or haplotype phasing?”

Instructor note. Encourage precise terminology: “overlap graph,” “repeat graph,” “string graph,” “homopolymer compression,” “haplotype bubble.” This synthesis stage consolidates algorithmic, biological, and computational aspects.

80–90 min – Reflection & Wrap-Up

Closing prompts:

- “What is one misconception you corrected today?”
- “Which assembly step (overlap detection, repeat handling, phasing, scaffolding) is still unclear?”
- “What tool or concept would you like a demonstration of in the next session?”

Instructor note. Reflection helps identify common pain points (phasing, Hi-C interpretation, repeat resolution). Use student feedback to calibrate future pacing and choose topics for deeper walkthroughs.