90-Minute Discussion Session Plan
Lecture 2: Sequencing & k-mers

Course: BINF301 — Computational Biology
Instructor: Tom Michoel
Date: 19/01/2026

0-10 min — Warm-Up

- Icebreaker: Which sequencing technology (Sanger, Illumina, PacBio, Nanopore) are you most familiar with?
- Starter prompt: What surprised you most about differences between sequencing generations?

Instructor note. Goal: lower anxiety barriers, establish conversational readiness, surface prior knowledge, and encourage all students to speak once very early.

10-25 min — Think-Pair-Share

Prompt: Based on the lecture slides, what do you think is the main limitation of each sequencing technology (1st, 2nd, 3rd generation)?

Follow-up: How do quality scores influence downstream analysis?

Instructor note. Students should identify: • Sanger: low throughput. • Illumina: short reads, drop in 3' quality. • PacBio/Nanopore: higher error rates (raw), cost. Quality scores → trimming → fewer spurious k-mers.

25-45 min — Structured Group Discussion

Roles:

- **Summarizer:** Recap FASTA vs FASTQ formats and Phred quality scores.
- **Questioner:** Bring one question about basecalling or sequencing error models.
- **Connector:** Link k-mers to sequencing output and quality issues.

Starter question: Why do sequencing errors inflate the number of unique k-mers?

Instructor note. Errors → new erroneous sequences → appear once → left-side tail of k-mer frequency plot → overestimates genome size.

Break (15 min)

45-65 min — Applied Exercises Block

Students work collaboratively on structured exercises, see handout.

80-90 min — Reflection & Wrap-Up

- One takeaway from today?
- One question to revisit next time?

Instructor note. Useful to collect feedback for pacing, confusion points, and future topics.