

60-Minute Discussion Session Plan

Lecture 7: Searching Genomes and Genome Indexing

Course: BINF301 – Computational Biology

Instructor: Tom Michoel

Date: 4/2/2026

Created with Copilot

0-5 min – Warm-Up

Prompts:

- “When you use Ctrl+F or grep, what do you think happens computationally?”
- “Why might naïve string matching fail for gigabase-scale genomes?”
- “Which concept from the pre-read seemed most confusing: Boyer–Moore, suffix arrays, or FM-index?”

5–20 min – Guided Concept Walkthrough

Purpose: build shared understanding before group work.

Topics to revisit:

- Why naïve pattern search is slow (Slides 4–6).
- Boyer–Moore logic: bad-character + good-suffix skipping (Slides 9–12).
- Why pattern preprocessing alone is insufficient for large-scale searching.
- Why indexing the *text* (genome) matters (Slides 17–18).

Guiding Questions:

- “How do Boyer–Moore’s rules avoid redundant comparisons?”
- “Why is the pattern scanned from right to left in Boyer–Moore?”
- “Why is text indexing crucial for read mapping and multi-query workloads?”

20–38 min — Structured Small-Group Discussion (Rotating Roles)

Students form groups of 3. Roles rotate every 6–7 minutes.

Roles

- **Summarizer:** explains how k -mer tables and hash tables provide fast fixed-length searches (Slides 19–23, 28–29).
- **Questioner:** asks about tradeoffs in suffix trees, suffix arrays, and FM-index (Slides 32–52).
- **Connector:** links indexing structures to real tools (mappers, aligners, search engines).

Starter Question

Why do we need different indexing structures for fixed-length and variable-length pattern searches?

40–50 min — Mini Applied Exercise Block

See handout.

50–60 min – Synthesis Discussion

Prompts:

- “If you needed to index the human genome on a laptop, which structure would you choose and why?”
- “How does the FM-index achieve both small size and fast lookup?”
- “What characteristics of genomes make indexing harder than indexing typical English text?”