Syllabus: CS 5124
Algorithms in Bioinformatics
Fall, 2003

Instructor: Lenwood Heath

- **Office:** 2160J Torgersen Hall
- **Office Hours:** 9:30–11 AM Tuesdays and Thursdays; 9-10:30AM Wednesdays
- **Email:** heath@vt.edu

Graduate Teaching Assistant: Ranjit Randhawa

- **Office:** 133 McBryde Hall
- **Office Hours:** To be announced on the course web site
- **Email:** rrandhaw@vt.edu

Class Meets: McBryde 219, 8–9:15 AM Tuesdays and Thursdays

Exams

<table>
<thead>
<tr>
<th>Exam</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midterm Exam</td>
<td>Tuesday, October 21, 8–9:15 AM</td>
</tr>
<tr>
<td>Final Exam</td>
<td>Wednesday, December 17, 4:25-6:25 PM</td>
</tr>
</tbody>
</table>

Index Number: 96345

Prerequisites:

- Data Structures (CS 2604) required
- CS 4104, Data and Algorithm Analysis, highly desirable
- **Corequisite:** PPWS 5314 — Biological Paradigms for Bioinformatics — or equivalent coursework in genetics and molecular cell biology

Textbook:

On Reserve:
For current list, see class web site.

Description
This course emphasizes algorithms to solve problems found in biology, especially molecular biology. A variety of current problems in computational molecular biology will be introduced, investigated, analyzed for computational complexity, and solved with efficient algorithms, when feasible. A number of such problems will be shown to be NP-complete or other evidence of their difficulty will be presented.

Grading Policy
Grading for the course is on a 1000-point scale, with the points distributed as follows:

Homework assignments: 12 at about 50 points each 600
Midterm exam: October 21, 8-9:15AM 150
Final exam: December 17, 4:25-6:25 PM 250

A typical homework assignment consists of 2 or 3 problems or exercises, posted on the web site. All homework must be prepared with \LaTeX or other word processing system and submitted as a stapled printout to a box outside the instructor’s office (2160J Torgersen Hall). Homework is due at 4:00 PM on the due date (see course calendar). No late homework will be accepted.

Ethics
The Honor Code applies. All work submitted must be the student’s own work. Students may solicit help only from the instructor or the GTA.

Announcement
If any student needs special accommodations because of a disability, please contact the instructor during the first week of classes.
## Intended Topics and Approximate Order

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>TOPIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course overview</td>
<td>Course overview</td>
</tr>
<tr>
<td>Gusfield 1</td>
<td>Exact matching: first algorithms</td>
</tr>
<tr>
<td>Gusfield 2.1–2.3</td>
<td>Exact matching: classical algorithms</td>
</tr>
<tr>
<td>Gusfield 3</td>
<td>Exact matching: a deeper look</td>
</tr>
<tr>
<td>Gusfield 5</td>
<td>Suffix trees</td>
</tr>
<tr>
<td>Gusfield 6.1–6.2</td>
<td>Linear-time construction of suffix trees</td>
</tr>
<tr>
<td>Gusfield 7</td>
<td>Selected applications of suffix trees</td>
</tr>
<tr>
<td>Gusfield 10</td>
<td>The importance of sequence comparison in molecular biology</td>
</tr>
<tr>
<td>Gusfield 11</td>
<td>Core string edits, alignments, and dynamic programming</td>
</tr>
<tr>
<td>Gusfield 14; Durbin, <em>et al.</em>, 6</td>
<td>Multiple string comparison and multiple sequence alignment</td>
</tr>
<tr>
<td>Papers from the literature; Durbin, <em>et al.</em>, 2.2, 2.7–2.8</td>
<td>Probability in bioinformatics; the statistical basis for scoring matrices; PAM and BLOSUM matrices</td>
</tr>
<tr>
<td>Gusfield 15; Durbin, <em>et al.</em>, 2.3–2.6</td>
<td>Sequence databases and searching — BLAST, PSI-BLAST, and FASTA</td>
</tr>
<tr>
<td>Durbin, <em>et al.</em>, 3–5</td>
<td>Hidden Markov models in bioinformatics</td>
</tr>
<tr>
<td>Gusfield 17; Durbin, <em>et al.</em>, 7–8</td>
<td>Evolutionary or phylogenetic trees; survey of algorithms for constructing phylogenetic trees; bootstrapping</td>
</tr>
<tr>
<td>Gusfield 16</td>
<td>Selected sections on mapping and sequencing</td>
</tr>
</tbody>
</table>