# Certificate in Bioinformatics

**Title of Program:** Certificate in Bioinformatics

**Level (Masters/Ph.D.):** Graduate Certificate

Please Indicate: ______ Program ___X___ Certificate _______ Concentration _______ Track

**Description of certificate, concentration or degree program:**
Please attach a description of the new certificate or concentration. Attach Course Inventory Forms for each new or modified course included in the program. For new degree programs, please attach the SCHEV Program Proposal submission.

**Please list the contact person for this new certificate, concentration, track or program for incoming students:**
Certificate Coordinator, Dr. John Grefenstette, 703-993-8400, jgrefens@gmu.edu.

**Approval from other units:**
Please list those units outside of your own who may be affected by this new program. Each of these units must approve this change prior to its being submitted to the Graduate Council for approval.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Head of Unit’s Signature</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Submitted by: ____________________________ Email: ____________
Graduate Council approval: ________________ Date: ____________
Graduate Council representative: ________________ Date: ____________
Provost Office representative: ________________ Date: ____________
# Graduate Certificate in Bioinformatics

Table of Contents

1. Executive Summary 3
2. Description of Proposed Graduate Certificate 3
   2.1. Mission and Objective 3
      2.1.1 Mission 3
      2.1.2 Objectives 4
   2.2. Administrative Structure 4
   2.3. Admission Requirements 4
   2.4. Curriculum Requirements 4
   2.5. Relationship to Other GMU Programs 5
3. Justification for Proposed Graduate Certificate 5
   3.1. Student Demand 5
   3.2. Employer Demand for Graduates 6
   3.3. Comparison with Other Programs in the Region and Commonwealth 6
   3.4. Projected Enrollment 7
   3.5. Differential Tuition Rate 7
4. Resource Needs 7
   4.1. Faculty 7
   4.2. Equipment and Computing Environment 8
1. Executive Summary

Because of the enormous increase in the availability and utilization of biological data, the School of Computational Sciences (SCS) has developed a new graduate Certificate in Bioinformatics (BINF) to meet the needs of prospective students, area employers, and society at large. This Certificate is administered by the Bioinformatics and Computational Biology Program within the School of Computational Sciences. The BINF certificate requires students to complete a set of 15-credit hours of SCS graduate courses. Ideal candidates for this certificate are those who have a background in biological and computer sciences, and are either currently working in or planning to enter into the fields of biotechnology or bioinformatics. We believe that the BINF certificate will prove very attractive to students who are interested in advancing their career goals, but who may not have adequate time available to undertake a graduate degree program. The 15-credit certificate is based upon the set of core courses currently supporting the BINF M.S. and Ph.D. degree programs, along with a set of elective courses. Students completing the BINF certificate will receive the most up-to-date advanced bioinformatics education available in the region. Completion of the certificate will enhance the careers of those students who are already working in this area, and can also serve as a useful intermediate step towards later enrollment in the Bioinformatics and Computational Biology M.S. or Ph.D. degree program.

2. Description of Proposed Graduate Certificate

2.1 Mission and Objectives

2.1.1 Mission

The innovative Certificate in Bioinformatics proposed here addresses the growing national and regional demand for trained computational biologists. Local corporations expected to hire graduates of the program include Celera, TIGR, Human Genome Sciences, and Gene Logic, as well as many other firms in the region’s robust biotechnology community. The proposed degree combines a solid foundation in biotechnology with computational skills relevant to bioinformatics. The flexibility of the Certificate structure permits students to custom-design their curriculum under an advisor’s guidance, making the Certificate in Bioinformatics especially relevant for students employed in today’s diverse Northern Virginia high-technology workplace. The proposed Certificate is intended for:

- Students seeking advancement in their current bioinformatics career.
- Students with a background in biological science or computing who are planning to enter the field of bioinformatics.
- Students in the Bioinformatics M.S. program whose career plans change, and would benefit from the availability of a Certificate.

All courses are offered in the late afternoon or early evening to accommodate students with full-time employment outside the university.

GMU has already developed thriving, interdisciplinary M.S. and Ph.D. programs in Bioinformatics. The proposed new Certificate is based upon a similar set of academic principles. The new Certificate program includes a core of fundamental bioinformatics classes as well as a flexible set of elective courses covering a wide range of topics in molecular biology, computational biology and software development. The Certificate utilizes an extensive base of existing coursework associated with the SCS Program in Bioinformatics and Computational Biology.
2.1.2 Objectives

Bioinformatics refers to the application of information technology to the storage, retrieval and analysis of information about biological sequences, structures and functions. As such, bioinformatics is an inherently multidisciplinary specialty that requires a background in both biology and computing. The Certificate is designed to meet the challenge of preparing students with diverse backgrounds for the field of bioinformatics. At the time of completion, students should be able to:

- Evaluate quantitatively the performance of bioinformatics algorithms and tools.
- Analyze, visualize, and interpret biological data.
- Work collaboratively in interdisciplinary groups.

2.2 Administrative Structure

The Certificate will be administered by the Bioinformatics Program within SCS. For this certificate, there will be a Certificate Coordinator, Prof. John Grefenstette, who will direct the administration of the proposed certificate. The Coordinator will supervise the admission of students into the program, monitor student progress, advise the students, promote the certificate, and coordinate course offerings. The mechanics of the admission process (i.e., application processing, student record control, etc.) will be handled by the Bioinformatics Program Admissions Processors.

2.3 Admission Requirements

a. Applicants to the BINF graduate certificate program should hold a B.A. or B.S. degree in a discipline related to biological and/or computer science from an accredited university, with a minimum GPA of 3.000.

b. Applicants should have taken courses in molecular biology, computer science, calculus, and/or statistics. Applicants should possess working knowledge of a computer programming language.

c. Applicants should submit the followings materials:
   - A completed GMU graduate application
   - Official transcripts
   - Resume
   - Virginia Domicile Classification form
   - TOEFL scores if they are foreign nationals
   - A check for $60 payable to George Mason University

   Note that GRE scores and letters of recommendation are not required.

d. Students may not pursue this certificate concurrently with any other graduate degree program or certificate program offered by SCS because this certificate program will charge students at a differential (premium) tuition rate. However, students enrolled in academic programs outside SCS may enroll in this certificate program concurrently.

Address procedural questions to the Certificate Coordinator, Dr. John Grefenstette, 703-993-8400, jgrefens@gmu.edu.

2.4 Curriculum Requirements
The Bioinformatics (BINF) Certificate requires a total of 15 credit hours, or five 3-credit courses. Students are required to take three core courses, plus a fifth course selected from the list of electives indicated below.

**Required Core Courses:** (all are required)
- BINF 630 – Bioinformatics Methods
- BINF 631 – Molecular Cell Biology for Bioinformatics
- BINF 634 – Bioinformatics Programming

**Elective Courses:** (choose two of the following or other courses approved by the coordinator)
- BINF 633 – Molecular Biotechnology
- BINF 636 – Microarray Methodology and Analysis
- BINF 639 – Biometrics
- BINF 730 – Biological Sequence Analysis
- BINF 731 – Protein Structure Analysis
- BINF 732 – Genomics
- BINF 733 – Gene Expression Analysis
- BINF 734 – Advanced Bioinformatics Programming
- BINF 739 – Topics in Bioinformatics

2.5 Relationship to other GMU Programs

No comparable certificate program is currently being offered at GMU. The coursework for the certificate program will overlap possible coursework for the MS and PhD in Bioinformatics. Students may opt to later apply to these programs or other programs at GMU if they meet the admissions criteria defined elsewhere.

3. Justification for Proposed Graduate Certificate

3.1 Student Demand

There is evidence indicating strong student demand for the proposed Certificate. The current M. S. Program in Bioinformatics has generated a high level of interest from potential students. In the last two years, over 250 inquiries have been received about the MS in Bioinformatics. There are currently over 50 students enrolled in the MS Program. There are also 15 students enrolled as non-degree seeking students. We expect that the Certificate Bioinformatics will attract a number of students greater than the set of non-degree students. Many of the new students that would enroll in the Certificate Program will be members of the local workforce who are interested in career advancement, but are perhaps not inclined to make the commitment to a M.S. or Ph.D. program at this time.

New GMU graduates with bachelor’s degrees in undergraduate majors such as computer science and biology, or those with an undergraduate minor in bioinformatics (within CAS), may also be potentially interested in the degree proposed here due to its interdisciplinary nature. We expect a significant number of these students to apply to the new program.

Students will be recruited via the SCS website. SCS will also place advertisements in journals and newspapers. Brochures will be prepared and sent to local biotechnology firms, including Celera, TIGR, Human Genome Sciences, GeneLogic, SAIC, and others. Brochures will also be sent to federal government laboratories and agencies including NIH, USDA, EPA, and FBI. Students will also be recruited at Open House events held at GMU.
3.2 Employer Demand for Graduates

The post-genome biology will be dominated by computational approaches to handling data as well as to computing new relationships within the data. It is anticipated that the computer will become the single most important piece of equipment in the biology lab, and the next generation of biology researchers will need to be trained in basic bioinformatics approaches.

A study for the National Academy of Science found that between 1996 and 1997 the number of distinct bioinformatics positions advertised nationally rose 68.6 percent, while the number of qualified graduates from formal or informal bioinformatics training programs were sufficient to fill only 15 percent of those vacancies. The difficulties associated with filling bioinformatics positions is reflected in the salaries: For the same time period the starting salary for bioinformatics M.S. averaged over $65,000 while the average for the life sciences was about half that figure. The demand continues to grow, with some projections suggesting that the bioinformatics industry might be $2.5 billion by 2005, a 12-fold increase from today's level, requiring as many as 20,000 additional trained workers.

These conclusions are supported by several additional reports, including one from the National Research Council, which states that the employment market for bioinformatics and information technology in general is "tight and likely to remain so for the immediate future," with "specialists in bioinformatics in great demand relative to supply." In 1999, the number of people working in the biotechnology industry in the U.S. was estimated to be 153,000, a 9 percent increase over the previous year. A joint report from the National Institutes of Health and the National Science Foundation states that "students who are not trained in integrated, multidisciplinary research are at a disadvantage." The NIH has recognized the critical need for training in bioinformatics to support the continued explosive growth in biotechnology. Furthermore, NIH has placed great emphasis on bioinformatics and has included several Bioinformatics and Computational Biology Initiatives as part of the NIH Roadmap.

Regionally, the demand for this degree is quite high. GMU is close to the NIH, the USDA, and the EPA, and each agency is represented by students in the current graduate program. Additionally, the Northern Virginia and Washington, D.C. metropolitan areas host one of the largest concentrations of biotechnology companies in the USA. All of these activities create a strong regional demand for workers with advanced training in Bioinformatics.

Our belief is that the proposed Certificate program in Bioinformatics along with the existing M.S. and Ph.D Programs completely address these needs, and will prove to be a highly competitive and demanded program.

3.3 Comparison with Other Programs in the Region and the Commonwealth

While the demand for bioinformatics professionals in this region and nationwide is relatively high and on the rise, most professionals receive education and training through formal degree programs, which require significant commitment of one’s time and resources. Little is available in post-graduate training beyond the traditional degree programs. The proposed graduate certificate therefore serves a special niche in meeting the needs of bioinformatics professionals in this technological area, and potentially for the nation.

There is no non-degree certification in the region that would somewhat be related to the proposed certificate. In non-official discussions we heard that Virginia Tech might be developing a certification program for medical applications.

We performed a survey on the Internet to identify programs in the Nation similar to the one we propose here. We found no existing programs in the Washington DC area. However, we found the Duke University, Stanford University, Leigh University, University of Pennsylvania have already pioneered with certificate degrees somehow similar to the one proposed here.

3.4 Projected Enrollment

With the expectation that the certificate will be approved and in place for Fall 2005, the enrollment in the initial year will likely be low, but the number will increase quickly when the news about the availability of the certificate spread throughout the local corporate community. Below are some projected figures for the number of students enrolling in the certificate program:

<table>
<thead>
<tr>
<th></th>
<th>Fall 2005</th>
<th>Fall 2006</th>
<th>Fall 2007</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>

In the long run, we expect that the student enrollment would stabilize around 16 students. However, we can accommodate higher numbers of students. We expect that students enrolling in the certificate will be primarily part-time students, and therefore they will take only one to two courses per semester. Consequently, on average, a student will complete the certificate program in about two years.

3.5 Differential Tuition Rate

The BINF graduate certificate will charge students at a differential (premium) tuition rate, with an additional $100 per credit hour added to the standard GMU graduate tuition rate for students who enroll in this certificate program (regardless of in-state or out-of-state status). Based on standard GMU tuition for the AY 2004-2005, in-state graduate students in the certificate program will therefore pay $345/credit hour and out-state students will pay $723/credit hour. When our tuition is compared with private institutions in the area, our rates are quite competitive. For instance, graduate students at George Washington University (GWU) are charged $810 per credit hour, and in Georgetown University, the rate is $993 per credit hour. The differential tuition will be used to fund continuing improvements in the SCS computational facilities used to support the certificate program.

4. Resource Needs

4.1 Faculty

The Program in Bioinformatics and Computational Biology in the School of Computational Sciences has a faculty body that would like to be involved in this certificate instruction. These primary faculties supporting the certificate and their respective courses are:

Prof. Blackwell
4.2 Equipment and Laboratory Environment

The Program in Bioinformatics has experimental laboratories, electronic classrooms, computer laboratories at their disposal that are routinely used in classes. There are also drop-in computer labs that students can use at their convenience.

APPENDIX – Course Descriptions

- **BINF 630 Bioinformatics Methods (3:3:0).** Prerequisites: Graduate standing or permission of instructor. Introduction to bioinformatics methods and tools for pairwise sequence comparison, multiple sequence alignment, phylogenetic analysis, protein structure prediction and comparison, database similarity searches, and discovery of conserved patterns in protein sequence and structures.

- **BINF 631 Molecular Cell Biology for Bioinformatics (3:3:0).** Prerequisites: Undergraduate background in biochemistry or cell biology, or permission of instructor. Intensive review of aspects of biochemistry, molecular biology, and cell biology necessary to begin research in bioinformatics. Topics include cell structure and cell cycle; DNA replication, transcription, and translation; molecular structure of genes and chromosomes.

- **BINF 633 Molecular Biotechnology (3:3:0).** Prerequisites: Graduate standing or permission of instructor. A laboratory intensive course introducing the theory and practice of modern biotechnology. Includes study of recombinant DNA, gene expression, and genetic analysis and associated methods. Laboratory exercises change to reflect the more recent advances in the field.

- **BINF 634 Bioinformatics Programming (3:3:0).** Prerequisites: Graduate standing and computer programming experience or permission of instructor. Data representation, control structures, file input/output, subroutines, regular expressions, debugging, introduction to relational databases. An emphasis on bioinformatics applications including DNA sequence analysis, parsing FASTA and GenBank files, processing BLAST output files, SQL or equivalent query language.

- **BINF 636 Microarray Methodology and Analysis (3:3:0).** Prerequisite: BINF 633 or permission of instructor. Introduces the theory and practice of genome analysis, including the genetics, biochemistry, and tools for analysis of global gene expression, as well as the detection and quantification of genes and gene products.

- **BINF 639 Biometrics (3:3:0).** Prerequisites: Programming experience (e.g. CSI 603 and 604) or permission of instructor. Introduction into methods for measuring humans. Topics include face recognition, speaker recognition, fingerprint recognition, shoeprint recognition, handwriting analysis, and other topics as time permits. Students will develop computer programs to perform many of these tasks.

- **BINF 730 Biological Sequence Analysis (3:3:0).** Prerequisites: BINF 702 or previous courses in programming, molecular biology, and probability, or permission of instructor. Fundamental methods for the analysis of nucleic acid and protein sequences, including pairwise alignment, multiple alignment, database
search methods, profile searches, and phylogenetic inference. Development of probabilistic tools, including hidden Markov models and optimization algorithms. Survey of current software tools.

- **BINF 731 Protein Structure Analysis (3:3:0)**. Prerequisite: Permission of instructor, or coursework in molecular biology, biochemistry, and computer programming. Computational methods for the analysis, classification and prediction of three-dimensional protein structures. The course covers theoretical approaches, techniques, and computational tools for protein structure analysis.

- **BINF 732 Genomics (3:3:0)**. Prerequisites: BINF 730 or previous courses in biology, numerical methods, and programming, or permission of instructor. A survey of computational tools and techniques used to study whole genomes. The biological basis of genome analysis algorithms will be explored. Lecture topics include genome mapping, comparative genomics, and functional genomics.

- **BINF 733 Gene Expression Analysis (3:3:0)**. Prerequisites: Programming experience and a course in molecular biology, or permission of instructor; S-Plus or Matlab experience may also be helpful. This course will focus on the analysis of gene expression data. Particular topics include: cluster analysis and visualization of expression data; inference of genetic regulatory networks; and theoretical models of genetic networks.

- **BINF 734 Advanced Bioinformatics Programming (3:3:0)**. Prerequisites: BINF 634 or permission of the instructor. Selected topics including algorithm design, complex data structures, object oriented programming, relational databases, designing modules, graphics programming, web programming. Students will complete a bioinformatics programming project.

- **BINF 739 Topics in Bioinformatics (3:3:0)**. Prerequisite: Permission of instructor. Selected topics in bioinformatics not covered in fixed-content bioinformatics courses. May be repeated for credit as needed.