Programming in Bioinformatics 2016/2017

Code: 42401
ECTS Credits: 6

<table>
<thead>
<tr>
<th>Degree</th>
<th>Type</th>
<th>Year</th>
<th>Semester</th>
</tr>
</thead>
<tbody>
<tr>
<td>4313473 Bioinformatics</td>
<td>OB</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Contact

Name: Antonio Miguel Espinosa Morales
Email: AntonioMiguel.Espinosa@uab.cat

Teachers

Juan Carlos Moure Lopez
Raquel Egea Sánchez

Use of languages

Principal working language: english (eng)

Prerequisites

For the general development of the course, it is recommended to have a B2 level, or equivalent, of English language.

For this module, it is very recommended to have basic notions of computer usage in Linux, knowledge of common user tools and basic statistics.

Objectives and Contextualisation

General objectives of this module are the application of the core tools and basic techniques for development in this area of knowledge. Provide skills to successfully assume the adaptation to changing technologies and new paradigms emerging in this interdisciplinary field.

Skills

- Design and apply scientific methodology in resolving problems.
- Identify the biocomputing needs of research centres and companies in the biotechnology and biomedicine sectors.
- Possess and understand knowledge that provides a basis or opportunity for originality in the development and/or application of ideas, often in a research context.
- Student should possess the learning skills that enable them to continue studying in a way that is largely student led or independent.
- Use operating systems, programs and tools in common use in biocomputing and be able to manage high performance computing platforms, programming languages and biocomputing analysis.
- Work individually and as part of a team in an international and multidisciplinary context.

Learning outcomes

1. Design, analyse and evaluate the performance of parallel infrastructures and large volumes of data.
2. Design and apply scientific methodology in resolving problems.
3. Identify the advantage and limitations of biocomputing and the importance of applying new computer technology in omic research.
4. Manage parallel platforms and biocomputing databases according to needs.
5. Master programming language in R and Perl Biocomputing.
6. Possess and understand knowledge that provides a basis or opportunity for originality in the development and/or application of ideas, often in a research context.
7. Student should possess the learning skills that enable them to continue studying in a way that is largely student led or independent.
8. Use algorithms and statistical calculation techniques to manage large volumes of data.
9. Use the main molecular databases, the main standard formats for molecular data and integrate data from different sources.
10. Work individually and as part of a team in an international and multidisciplinary context.

Content

Introduction to Module
Professor Juan Carlos Moure

Computational Thinking
Professor Juan Carlos Moure

- Why computers and biology?
- Computational workflows
- Computer Programs
- Computer Architecture
- Programming concepts
- Algorithms and data structures

Programming languages: R
Professor Toni Espinosa

- R programming
- Bioconductor

Linux (commands and shell scripting)
Professor Toni Espinosa

- Basic linux: basic commands and editors, user management, software management, file system
- Shell scripting: terminal tools, variables and execution environment, examples and exercises

Programming Languages and Algorithms
Professors Toni Espinosa and Juan Carlos Moure

- Overview of programming languages in Bioinformatics
- Data types and contexts
- Data manipulations and operations
- Input/Output
- Code reutilization: modules and subroutines
- Regular expressions
- Third party code integration
Database management with MySQL

Professors Toni Espinosa

- Introduction to databases
- DBMS structure, MySQL
- Creation of biological Databases
- Query analysis of biological databases

Methodology

The methodology will combine classroom work, supervised problem solving in class, unsupervised work in the computing lab, homework from recommended readings and independent study student. It will use the virtual platform and asked for papers related to the thematic blocks.

Activities

<table>
<thead>
<tr>
<th>Title</th>
<th>Hours</th>
<th>ECTS</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type: Directed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classroom work</td>
<td>20</td>
<td>0.8</td>
<td>2, 5, 7, 9, 10</td>
</tr>
<tr>
<td>Problem solving (in class)</td>
<td>14</td>
<td>0.56</td>
<td></td>
</tr>
<tr>
<td>Work in the computing lab</td>
<td>12</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td><strong>Type: Supervised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performing lab work from recommending reading</td>
<td>15</td>
<td>0.6</td>
<td></td>
</tr>
<tr>
<td><strong>Type: Autonomous</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular general work on the deliverables definition and materials given</td>
<td>83</td>
<td>3.32</td>
<td>2, 4, 5, 6, 8, 9, 10</td>
</tr>
</tbody>
</table>

Evaluation

The methodology will combine classroom work, problem solving in the classroom, unsupervised work done in the computing lab and individual work from recommended readings. It will make use of the virtual platform and will make references to selected publications related to the thematic blocks.

Evaluation activities

<table>
<thead>
<tr>
<th>Title</th>
<th>Weighting</th>
<th>Hours</th>
<th>ECTS</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation of work done during the module, presented by the user</td>
<td>10%</td>
<td>1</td>
<td>0.04</td>
<td>2, 5, 7, 9, 10</td>
</tr>
<tr>
<td>Final Exam</td>
<td>10%</td>
<td>1</td>
<td>0.04</td>
<td>2, 10</td>
</tr>
<tr>
<td>Individual theoretical and practical tests</td>
<td>50%</td>
<td>2</td>
<td>0.08</td>
<td>1, 2, 3, 4, 10</td>
</tr>
<tr>
<td>Laboratory work, possibly in groups</td>
<td>30%</td>
<td>2</td>
<td>0.08</td>
<td>2, 4, 5, 6, 8, 9, 10</td>
</tr>
</tbody>
</table>
Bibliography

- Bell, Charles; Kindahl, Mats; Thalmann, Lars. "MySQL High Availability". O'Reilly, 2010.
- Bessant, C., Shadford, I., Oakley, D. "Building Bioinformatics Solutions with Perl, R and MySQL", Oxford University Press, 2009
- Mäkinen et al., Genome-Scale Algorithm Design: Biological Sequence Analysis in the Era of High-Throughput Sequencing. Cambridge Univ. Press, 2015.

Recommended websites

- http://mscbioinformatics.uab.cat
- https://cv.uab.cat

Search for bioinformatics and computer science topics in UAB library e-book resources:

- http://www.uab.cat/biblioteques/trobador
- http://sfx.cbuc.cat/uab/azbook
- http://pagines.uab.cat/bctdigital/
- http://cataleg.uab.cat/search~S1+cat/l?SEARCH=(02.034)57

Linux manual: http://materials.cv.uoc.edu/continguts/PID_00148368/index.html