

**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR**  
**DEPARTAMENTO DE BIOLOGIA MOLECULAR E BIOTECNOLOGIA**

**Disciplina:** R para Ciências da Vida

**Período Letivo:** 2025/2

**Professor Responsável:** DIEGO BONATTO

**Sigla:** BCM13065 **Créditos:** 4

### Súmula da disciplina

Fundamentos de R: programas e interfaces; operadores, variáveis, funções e diretórios; pacotes para manipulação e criação de vetores, matrizes, tabelas e listas; tipos de dados biológicos e suas entradas; boas práticas de computação em R: ambiente, scripts e projetos; pacotes para análises de dados de larga escala (genômica, transcritômica e proteômica); pacotes para análises estatísticas; pacotes para a criação de gráficos e representação visual de dados.

**Pré-requisitos:** não há a necessidade de conhecimentos prévios em R ou qualquer outra linguagem de programação.

### Objetivos

A disciplina tem como objetivo central proporcionar uma compreensão abrangente da programação em R, bem como as suas ferramentas e aplicações focadas para as Ciências da Vida. Para tanto, a disciplina busca familiarizar os discentes com a linguagem e a programação R para dados biológicos. Nesse sentido, o objetivo central da disciplina será alcançado: (I) pela compreensão dos diferentes componentes da programação em R, como operadores e funções, para análise de dados biológicos, (ii) pela escolha da forma de entrada de dados biológicos uni e multidimensionais, (iii) pelo processamento dos dados biológicos e uso de pacotes específicos para a descoberta de conhecimento biológico, (iv) pelo emprego específico de pacotes gráficos para a representação visual dos resultados gerados e (v) pelo emprego de boas práticas de programação.

### Conteúdo programático e cronograma

Aula (Data)	Título	Conteúdo
1 (10/11/2025)	Introdução à disciplina; Histórico do R e a sua importância para a análise de dados biológicos (presencial e virtual)	Programas para análises estatísticas de dados biológicos; o programa "S". O desenvolvimento do R como plataforma para a resolução e análise de dados biológicos. Por que usar o R? Comparação com outras linguagens de programação (Python, C/C++).
2 (11/11/2025)	Procedimentos iniciais com o R: interfaces e atualizações (virtual)	Escolha de interfaces para o R: IDEs e GUIs. Procedimentos para a instalação de pacotes e a sua manutenção. Resolvendo problemas de dependências de pacotes (terminal ou console do R).
3 (12/11/2025)	Boas práticas de programação em R (virtual)	Uso de comentários e registro de linhas de comandos/scripts. Criação de arquivos do tipo R script, Notebook e Markdown.
4 (17/11/2025)	Pacotes do R para a análise de dados biológicos (virtual)	O sistema de arquivos CRAN. Bioconductor e GitHub como fontes de pacotes para análises biológicas. Cuidados específicos para instalação de pacotes no R (dependências e atualizações). Pacotes usados para análises de dados de larga escala, estatísticos e taxonomia/filogenia.
5 (18/11/2025)	Objetos do R (virtual)	O que são vetores, fatores, matrizes, arranjos, tabelas e listas. A importância de cada objeto para análises biológicas.
6 (19/11/2025)	Mineração e prospecção de bancos de dados biológicos (virtual)	Tipos de pacotes para data mining e knowledge discovery. Integração de ferramentas e pipelines para busca de dados biológicos.
7 (24/11/2025)	Processamento de dados biológicos – partes 1 e 2 (virtual)	Ferramentas do R para a manipulação de tabelas e listas. Remoção de dados e filtragem. Fusão de dados multidimensionais e a sua representação. Uso do "tidyverse" para o processamento de dados. Dplyr, Plyr, Magrittr, data.table e tidyr. Operadores e "pipes" para otimização dos dados biológicos.
8 (25/11/2025)	Gráficos e visualização de dados – partes 1 e 2 (virtual)	Funções nativas do R para a representação de dados biológicos (gráficos de barras, histogramas, dispersão de pontos, boxplots). O pacote ggplot2 e as suas aplicações gráficas. Otimização de cores para representação gráfica de dados.
9 (26/11/2025)	Interfaces com outras linguagens e plataformas (presencial e virtual)	Interfaces com Python. Uso do Latex para a geração de documentos.

## **Metodologias de ensino**

As vídeo-aulas serão ministradas na plataforma Moodle da UFRGS de forma síncrona e gravada para posterior disponibilização do conteúdo de forma assíncrona pela disponibilização de endereços de compartilhamento de arquivos na nuvem para posterior acesso. Todos os materiais necessários para o aprendizado, como capítulos de livros, artigos e textos da área, exercícios e R scripts serão disponibilizados na plataforma GitHub. Os R scripts gerados pelos discentes serão disponibilizados na plataforma Google Drive.

Teórica: 60 horas

## **Experiências de aprendizagem**

Sendo esta uma disciplina de caráter teórico-prático, os alunos terão contato com os conceitos de biologia sintética e o seu ferramental, propiciando um embasamento teórico completo para a resolução de problemas e contextos comumente encontrados na biologia sintética e na biotecnologia como um todo.

## **Critérios de Avaliação**

O sistema de avaliação constará na montagem de scripts em R e a sua disponibilização na plataforma Google Drive (peso dez).

Será considerado aprovado na disciplina o aluno que obtiver um nível de aproveitamento igual ou superior a 60% (sessenta por cento).

Àqueles alunos que não obtiverem um nível de aproveitamento igual ou superior a 60% (sessenta por cento) será facultada a recuperação abrangendo os conteúdos vistos durante o semestre .

A nota da recuperação substituirá a nota correspondente da avaliação regular.

O conceito final será atribuído com base na média aritmética das notas obtidas nos cinco trabalhos individuais e na avaliação (TI e AV), considerado o resultado da recuperação, quando for o caso. A fórmula para cálculo da média aritmética é a seguinte:  $MA = \text{Somatório das notas TI + AV} / 2$

As notas serão convertidas em conceitos conforme a tabela abaixo: e seguirá a seguinte distribuição:

A - 90% a 100%

B - 75% a 89%

C - 60% a 74%

D - aproveitamento inferior a 60%

FF - frequência insuficiente à disciplina (inferior a 75%)

Será facultada uma prova de recuperação abrangendo os conteúdos vistos durante o semestre. Os resultados serão divulgados até dois dias após a avaliação.

## **Bibliografia**

### **Básica Essencial**

Vinhetas e resumos dos principais pacotes para análise de dados biológicos, geração e manipulação de estrutura de dados e para a representação gráfica a serem fornecidos pelo docente ao longo da disciplina.