

# 序列处理软件 用户使用手册

## 目录

1. 引言
2. 系统需求与安装
3. 启动应用程序
4. 界面总览
5. 功能模块详解
  - 5.1 模块 1：保守区分析
  - 5.2 模块 2：序列导出
  - 5.3 模块 3：差异度计算
  - 5.4 模块 4：同一率计算
  - 5.5 模块 5：序列长度计算
  - 5.6 模块 6-1：从 NCBI 获取全基因组
  - 5.7 模块 6-2：从 NCBI 获取 CDS
  - 5.8 模块 7：序列优化 (SeqRefine)
  - 5.9 模块 8：引物设计
  - 5.10 模块 9：一/二维码可视化
6. 常用操作
  - 6.1 语言切换
  - 6.2 模块切换
  - 6.3 文件与文件夹浏览
7. 输出结果说明
8. 故障排查
9. 技术支持与反馈

---

## 1. 引言

欢迎使用“序列处理软件”——一款集保守区分析、序列导出、差异度与同一率计算、序列长度统计、NCBI 数据抓取、序列优化、引物设计及一/二维码生成于一体的可视化工具。

本手册将详细介绍软件的安装、界面操作、各功能模块的使用方法与注意事项，帮助用户快速上手并高效完成生物信息学数据处理任务。

## 2. 系统需求与安装

1. **操作系统：** Windows 10/11, macOS 10.15 及以上，Linux（支持 PyQt5）
2. **Python 版本：** Python 3.8 及以上
3. **依赖库：**

- PyQt5
- Biopython
- numpy
- pandas
- scikit-learn
- seaborn
- matplotlib
- python-docx
- qrcode
- pillow

4. **安装步骤：**

```
pip install PyQt5 biopython numpy pandas scikit-learn seaborn  
matplotlib python-docx qrcode pillow
```

5. **下载程序：** 将 `module9.py` 及相关图标文件放置于同一文件夹。确保 软件图标.png 在程序可访问路径中。

## 3. 启动应用程序

1. 打开终端（命令提示符）。
2. 切换到程序目录：

```
cd /path/to/your/software
```

3. 运行主程序：

```
python module9.py
```

4. 程序启动后，将出现主窗口，准备接受用户操作。

## 4. 界面总览

主窗口包含以下核心区域：

- **顶部工具栏：**语言选择下拉框 + 功能模块选择下拉框
- **中部（StackedWidget）：**根据所选模块动态展示对应的输入面板
- **底部：**各模块的“运行”按钮，用于触发对应功能
- **样式：**渐变背景、统一字体（Arial）、富有层次的控制件配色方案。

## 5. 功能模块详解

以下各节依次介绍模块功能、输入参数、操作流程与输出结果。

### 5.1 模块 1：保守区分析

**功能概述：**提取多条 FASTA 序列中的高保守区段。

**输入参数：**

- **输入文件夹：**FASTA 文件（.fasta/.fas）所在目录
- **输出文件夹：**分析结果保存目录
- **最小保守区长度（bp）**
- **最大保守区长度（bp）**
- **缺口阈值（0.0–1.0，默认 0.1）：**过滤含 gap 过多的列
- **保守性阈值（0.0–1.0，默认 0.8）：**决定是否保留某列作为保守位点

**操作步骤：**

1. 在模块 1 面板中，点击“Browse...”选择输入 FASTA 文件夹。
2. 点击“Browse...”选择输出结果文件夹。
3. 在对应文本框中输入最小/最大长度及阈值。
4. 点击底部“运行模块 1”按钮，软件自动完成阈值调整、缺口处理、SVM 预测并导出保守片段。

**输出结果：**每个输入 FASTA 文件对应生成 conserved\_sequences\_<filename>.txt，内含：

- 序列片段编号、文件名、实际起始位置、验证结果
- 平均交叉验证准确率报告行

### 5.2 模块 2：序列导出

**功能概述：**根据已知保守区分段，将测试序列与保守序列分割并导出。

#### 输入参数:

- 输入 FASTA 文件: 待分割的 `.txt` 格式 FASTA 数据
- 输出文件夹: 分段结果保存目录
- 拆分索引: 第  $k$  条序列前为测试序列, 后续为保守序列

#### 操作步骤:

1. 选择待处理 FASTA 文本文件。
2. 设置输出文件夹路径。
3. 输入拆分索引 (例如 **1** 表示第 1 条之前为测试序列)。
4. 点击“运行模块 2”, 自动生成多条 `_segment_*.txt` 文本文件。

#### 输出结果:

- 测试序列和保守序列分别按行拆分后保存, 可用于后续差异/同一率分析。

## 5.3 模块 3: 差异度计算

**功能概述:** 计算测试序列与物种序列的平均核苷酸差异率, 并生成分布细节与热图。

#### 输入参数:

- 待测试序列文件 (`.txt`)
- 物种序列文件夹 (包含 `_segment_*.txt`)
- 输出 Excel 文件: 差异计算结果保存路径 (`.xlsx`)

#### 操作步骤:

1. 选择测试序列文本文件。
2. 选择物种序列段文件夹。
3. 选择并输入输出 Excel 文件路径。
4. 点击“运行模块 3”, 软件自动计算:
  - 平均核苷酸级差异 (%)
  - 各阈值比例分布 ( $\text{Diff} \geq X \%$ )
  - 各碱基数阈值分布 ( $\text{Diff} \geq X \text{ bases}$ )
  - 导出 Excel 三个 Sheet: 平均差异、比例分布、数量分布。
5. 结果生成后自动绘制并保存两张热图 (`_prop_heatmap.svg/.png` 和 `_num_heatmap.svg/.png`)。

**输出结果:** Excel 报表与对应热图, 用于快速定位差异显著的序列段。

## 5.4 模块 4：同一率计算

**功能概述：**计算测试序列与物种序列的平均一致性，同差异度模块类似。

**输入参数：**同模块 3。

**操作步骤：**

1. 填写测试序列、物种序列路径。
2. 设置输出 Excel 文件名称。
3. 点击“运行模块 4”，自动执行算法：
  - 平均核苷酸级同一率 (%)
  - 各阈值物种比例 (Identity > X %)
  - 各碱基数阈值分布 (Identity >= X bases)
  - 导出三个 Sheet 并生成两张热图。

## 5.5 模块 5：序列长度计算

**功能概述：**统计 FASTA 文件中每条序列的长度。

**输入参数：**

- 输入 FASTA 文件 (.fasta/.fas)
- 输出 Excel 文件路径 (.xlsx)

**操作步骤：**

1. 选择输入 FASTA 文件。
2. 设置输出 Excel 文件路径。
3. 点击“运行模块 5”，软件读取每条序列并统计长度，导出表格。

## 5.6 模块 6-1：从 NCBI 获取全基因组

**功能概述：**批量根据 Accession 列表抓取 GenBank 数据库中的完整基因组。

**输入参数：**

- Accession 列表文件（每行一个 Accession）。
- 输出文件夹（FASTA 格式）。

**操作步骤：**

1. 选择 Accession 列表文本文件。

2. 选择输出文件夹。
3. 点击“运行模块 6-1”，程序调用 Entrez API 下载并保存每个 accession 的 `.fasta` 文件。

## 5.7 模块 6-2：从 NCBI 获取 CDS

**功能概述：**根据 Accession 列表抓取并提取 CDS 区序列。

**输入参数：**同 6-1。

**操作步骤：**

1. 选择 Accession 列表文件。
2. 选择输出目录。
3. 点击“运行模块 6-2”，程序下载 GenBank 格式并筛选 CDS 特征，输出 `_cds.txt` 文件。

## 5.8 模块 7：序列优化 (SeqRefine)

**功能概述：**去除简并碱基并批量修改 FASTA 标题。

**输入参数：**

- FASTA 文件夹。
- 是否修改标题（是/否）。
- 新标题文本。
- 是否按文件名重命名。

**操作步骤：**

1. 选择包含 `.fasta/.fas` 文件的目录。
2. 勾选“修改序列标题”，并输入新标题（可选）。
3. 勾选“按文件名重命名”，软件自动替换标题。
4. 点击“运行模块 7”，程序去除所有简并碱基并完成重命名。

## 5.9 模块 8：引物设计

**功能概述：**自动为每条序列生成前向/后向引物并计算 GC%、Tm。

**输入参数：**

- 输入 FASTA 文件。
- 输出 Excel 文件路径。

- 引物长度（默认为 18 bp）。

操作步骤：

1. 选择输入 `.fasta` 文件。
2. 设置输出 Excel 路径。
3. 输入引物长度。
4. 点击“运行模块 8”，程序解析 FASTA 并导出包含引物序列、GC% 和 Tm 的表格。

## 5.10 模块 9：一/二维码可视化

**功能概述：**将文本或 DOCX 文件内容转换为美化二维码，并将原序列碱基替换为可视化符号，生成 Word 文档及二维码图片。

输入参数：

- 输入 TXT / DOCX 文件。
- 输出 Word 文档（`.docx`）。

操作步骤：

1. 选择文本或 DOCX 文件。
2. 设置输出 Word 文件路径。
3. 点击“运行模块 9”，程序自动拆分文本、生成带圆角渐变色二维码 PNG、替换碱基并生成最终 Word 文档。

# 6. 常用操作

## 6.1 语言切换

- 顶部左侧下拉框可在 `English` 与 `中文` 之间切换。
- 切换后，界面所有文本即时更新。

## 6.2 模块切换

- 顶部中间下拉框列出模块 1-9。
- 选择后，中部面板自动切换到对应模块输入界面。

## 6.3 文件与文件夹浏览

- 点击各模块面板中的“Browse...”按钮，弹出文件/文件夹选择对话框。
- 选择完成后，路径自动填入对应文本框。

## 7. 输出结果说明

- 所有 Excel、TXT 和 PNG/SVG 图片均保存在用户指定路径。
- 文件命名会依据模块与输入文件自动添加前缀/后缀。
- 热图：\_prop\_heatmap（比例热图）、\_num\_heatmap（数量热图）。

## 8. 故障排查

1. 依赖缺失：检查是否已安装所有第三方库。
2. 文件路径错误：确认目录或文件存在且有读写权限。
3. Excel 导出失败：尝试关闭 Excel 进程或更换输出文件名。
4. NCBI 下载异常：检查网络连接与 Entrez.email 设置。
5. GUI 无响应：重启程序或检查 Python 与 PyQt5 兼容性。

## 9. 技术支持与反馈

如有疑问或建议，请联系维护团队：

- 邮箱：gs2022@hnu.edu.cn

感谢您使用本软件，祝研究顺利！