

# strobealign-project 用户说明

## 项目简介

本项目实现了基于 strobemer 的 DNA 序列比对工具，支持参考序列与查询序列的快速比对，并可输出比对结果的 CSV 文件和分布可视化图片。适合生物信息学入门学习和小型基因组数据分析。

## 目录结构

```
strobealign-project/
├── src/
│   ├── strobealign.py      # 主算法与工具函数
│   └── utils.py            # 辅助工具
├── tests/
│   └── test_strobealign.py  # 单元测试与可视化
├── data/
│   ├── ref.fa              # 参考序列（FASTA格式）
│   └── query.fa            # 查询序列（FASTA格式）
├── results/                # 输出结果文件夹（可选）
└── README.md               # 用户说明
```

## 环境依赖

- Python 3.7 及以上
- [Biopython](#)
- matplotlib

安装依赖（建议使用虚拟环境）：

```
pip install biopython matplotlib
```

## 数据准备

请将你的参考序列和查询序列以FASTA格式放入 `data/` 文件夹，例如：

**ref.fa**

```
>ref1
ACGTACGTACGTACGTACGTACGTACGTACGTACGT
```

**query.fa**

```
>query1
ACGTACGTACGT
```

# 运行方法

---

## 1. 运行单元测试（推荐）

在项目根目录下执行：

```
python -m unittest tests/test_strobealign.py
```

测试会自动：

- 加载 data/ref.fa 和 data/query.fa
- 进行 strobemer 比对
- 输出比对结果到 CSV 文件
- 生成比对分布的 PNG 图片

每次运行都会生成带时间戳的唯一结果文件，例如：

```
ref_vs_query_20240605_153012_alignments.csv  
ref_vs_query_20240605_153012_distribution.png
```

## 2. 直接调用主函数

你也可以在 src/strobealign.py 中直接调用相关函数，进行自定义比对和输出。

---

## 输出说明

- **CSV 文件**  
包含所有比对结果，字段为：参考序列起点、查询序列起点、匹配长度。
- **PNG 图片**  
展示比对在参考序列上的分布，并带有详细中文批注（最大比对数、均值、总数等）。

---

## 常见问题

- **没有比对结果？**  
请检查 query.fa 和 ref.fa 是否有重叠区域，并适当调小参数（如 k、w\_min、w\_max、n）。
- **matplotlib 报错？**  
请确认已正确安装 matplotlib，并在虚拟环境下运行。
- **Webview 报错？**  
这不影响 Python 代码运行，只影响 VS Code 的可视化界面，可重启 VS Code 解决。

---

## 联系方式

如有问题或建议，请在本项目仓库提交 issue，或联系项目维护者。

---

## ## 项目简介

本项目实现了基于 **strobemer** 的 **DNA** 序列比对工具，支持参考序列与查询序列的快速比对，并可输出比对结果的 **CSV** 文件和分布可视化图片。适合生物信息学入门学习和小型基因组数据分析。

---

## ## 目录结构

```
strobealign-project/
├── src/
│   ├── strobealign.py    # 主算法与工具函数
│   └── utils.py          # 辅助工具
├── tests/
│   └── test_strobealign.py # 单元测试与可视化
├── data/
│   ├── ref.fa            # 参考序列 (FASTA格式)
│   └── query.fa          # 查询序列 (FASTA格式)
├── results/              # 输出结果文件夹 (可选)
└── README.md             # 用户说明
```

---

## ## 环境依赖

- Python 3.7 及以上
- [Biopython](https://biopython.org/)
- matplotlib

安装依赖（建议使用虚拟环境）：

```
```bash
pip install biopython matplotlib
```

# 数据准备

请将你的参考序列和查询序列以FASTA格式放入 `data/` 文件夹，例如：

## ref.fa

```
>ref1
ACGTACGTACGTACGTACGTACGTACGTACGTACGT
```

## query.fa

```
>query1
ACGTACGTACGT
```

# 运行方法

---

## 1. 运行单元测试（推荐）

在项目根目录下执行：

```
python -m unittest tests/test_strobealign.py
```

测试会自动：

- 加载 data/ref.fa 和 data/query.fa
- 进行 strobemer 比对
- 输出比对结果到 CSV 文件
- 生成比对分布的 PNG 图片

每次运行都会生成带时间戳的唯一结果文件，例如：

```
ref_vs_query_20240605_153012_alignments.csv  
ref_vs_query_20240605_153012_distribution.png
```

## 2. 直接调用主函数

你也可以在 src/strobealign.py 中直接调用相关函数，进行自定义比对和输出。

---

## 输出说明

- **CSV 文件**  
包含所有比对结果，字段为：参考序列起点、查询序列起点、匹配长度。
- **PNG 图片**  
展示比对在参考序列上的分布，并带有详细中文批注（最大比对数、均值、总数等）。

---

## 常见问题

- **没有比对结果？**  
请检查 query.fa 和 ref.fa 是否有重叠区域，并适当调小参数（如 k、w\_min、w\_max、n）。
- **matplotlib 报错？**  
请确认已正确安装 matplotlib，并在虚拟环境下运行。
- **Webview 报错？**  
这不影响 Python 代码运行，只影响 VS Code 的可视化界面，可重启 VS Code 解决。

---

## 联系方式

如有问题或建议，请在本项目仓库提交 issue，或联系项目维护者。

---