

用户手册：Minimap2 Simplified

📌 一、软件简介

本项目是对 Minimap2 核心比对流程的简化 Python 实现。通过 minimizer 索引、anchor 匹配、链化（chaining）和局部比对（Smith-Waterman），实现了从原始序列到比对输出的完整流程。

github链接：https://github.com/wangyukai585/bio2502_final_project

二、安装方法

克隆项目并进入目录

```
git clone https://github.com/your-name/minimap2_simplified.git
cd minimap2_simplified
```

安装依赖（推荐虚拟环境中执行）

```
pip install -r requirements.txt

pip install -e # 配置python环境
```

三、输入输出文件说明

输入文件

文件	格式	说明
ref.fa	FASTA	参考基因组序列
query.fa	FASTA	查询序列

输出文件

文件	格式	说明
*.csv	CSV	匹配 anchor 坐标点
*.paf	PAF	标准比对结果
图像窗口	PNG/GUI	anchor 分布图、最佳链分布图（matplotlib 显示）

四、使用方法

方式 1：运行示例数据（内置）

```
python examples/demo.py
```

会比对 toy 查询序列和参考序列，输出 anchor 图 + CSV + PAF。

方式 2：运行真实序列

```
python examples/fasta_run.py
```

请确保 examples/ 目录下已准备好 ref.fa 和 query.fa 。

五、参数说明（可修改）

参数名	默认值	模块	说明
k	5	minimizer.py	k-mer 长度
w	10	minimizer.py	滑动窗口大小
max_gap	50	chain.py	anchor 间最大跳跃容忍距离

如需修改，可在 fasta_run.py 调用函数处更改：

```
run_fasta_alignment("ref.fa", "query.fa", k=7, w=5)
```

六、输出结果解释

query_vs_ref.csv

```
ref_pos,query_pos
12345,10
12351,16
...
```

- 表示每个 anchor 匹配点在参考和查询中的位置

query_vs_ref.paf

```
query1  300 37  148 + ref1  4641652 515050  515206  63  156 60
```

字段说明如下：

字段序号	含义
1	查询 ID（query name）
2	查询长度
3-4	查询起止坐标（匹配区）
5	+/- 方向
6	参考 ID（ref name）
7	参考长度
8-9	参考起止坐标（匹配区）
10	匹配 anchor 数量
11	匹配区长度
12	Mapping quality

图像输出

程序会弹出窗口显示：

- 所有 anchor 散点图（表示全部匹配点）
- 最佳 chain 路径图（表示最终保留路径）

七、常见问题排查

问题	可能原因	解决方式
<code>FileNotFoundError</code>	输入路径错误	确保 <code>ref.fa</code> 和 <code>query.fa</code> 在指定位置
图像不显示中文	默认字体不支持	修改 <code>visualize.py</code> 中的字体设置
输出 <code>No anchors found...</code>	没有匹配的 minimizer	检查是否参考和查询相差太大或参数设置不当