

# 开 题 报 告

专业：生物信息学

班级：F1908004

学生姓名：刘澄、唐志轩、余文韬

指导老师：吴茂英

## 一、 选题：

实现在一组 DNA 序列中寻找转座子的精确匹配的算法。

## 二、 选题背景：

1. 转座子：转座子 (Transposon) 亦称为转座元件，跳跃子，是一类 DNA 序列，它们能够在基因组中通过转录或逆转录，在内切酶 (Nuclease) 的作用下，在其他基因座上出现。转座子可以分为 I 型转座子和 II 型转座子两类：前者又称为逆元件 (Retro element)，转座中间体为 RNA。该型转座子会先被转录为 RNA，然后该 RNA 被逆转录再次成为 DNA，才被插入到目标位点中；后者被称为不复制转座子，被内切酶切下后直接插入到目标位点中。
2. 转座子在生物 DNA 的主要组成部分（如人类基因组有大约 44-45% 的转座子），在研究生物的遗传变异中有重要作用。
3. 高通量测序方法的应用，给我们带来了大量的基因序列信息。通过基因组序列比对寻找某一生物基因组中含有的转座子，对研究该生物基因的遗传变异有着重要意义。

## 三、 研究现状：

实现基因组序列比对常有一下算法：

1. 简洁的哈希索引
2. 建立后缀树或后缀数组
3. 通过 Burrows-wheeler 变换压缩序列大小

## 四、 研究计划：

1. 对上述算法进行分析比较，选择一个可行性较大的算法进行实现。
2. 预实验：生成长随机 DNA 序列与短随机 DNA 序列比对。
3. 在某生物基因组（如大肠杆菌）中寻找已知转座子。

## 五、 参考文献：

[1]转座子. (2021. 2. 7). 维基百科. from  
<https://zh.m.wikipedia.org/wiki/%E8%BD%AC%E5%BA%A7%E5%AD%90>