



华中农业大学

HUAZHONG AGRICULTURAL UNIVERSITY

生物信息学进展 实验报告

ChIP-Seq 数据处理

学 院：信息学院

专 业：生物信息学

学生姓名：张子栋

学 号：2020317210101

湖北·武汉

2023 年 3 月 11 日

目录

1	实验材料	1
1.1	线虫参考基因组	1
1.1.1	WormBase	1
1.2	线虫 ChIP-Seq 数据	1
1.3	实验软件	1
1.3.1	BWA	1
1.3.2	Bowtie	2
2	实验步骤	3
3	实验结果	3
4	实验总结	3

1 实验材料

1.1 线虫参考基因组

1.1.1 WormBase

WormBase 是一个专门收录线虫的基因组信息的数据库，它支持使用线虫作为模式生物的研究者，提供了线虫的基因组序列、注释、变异、表达、互作等数据。WormBase 还包括一个子项目 WormBase ParaSite，它收录了其他线虫和扁形动物寄生虫的基因组信息。参考基因组是一个完整的线虫基因序列，基因组注释文件是对参考基因组中的基因、转录本、外显子等特征的描述。

通过 WormBase : Nematode Information Resource 网站获取线虫参考基因组。

1.2 线虫 ChIP-Seq 数据

来自于上次实验。

1.3 实验软件

1.3.1 BWA

BWA 是一种能够将差异度较小的序列比对到一个较大的参考基因组上的软件包。它由三个不同的算法组成：BWA-backtrack, BWA-SW 和 BWA-MEM。BWA-backtrack 适用于比对 Illumina 的序列，reads 长度最长能到 100 bp。BWA-SW 和 BWA-MEM 适用于比对长序列，支持的长度为 70-1 Mbp；同时支持剪接性比对。

BWA 的使用需要先对参考基因组建立索引，然后再进行比对。比对的结果是一个 SAM 格式的文件，可以用 samtools 进行后续的处理。

- 建立索引：`index ref.fa`
- 比对：`bwa mem ref.fa reads.fq > aln.sam`
- 查看帮助：`bwa -h`

1.3.2 Bowtie

Bowtie 是一个超快的，存储高效的短序列片段比对程序。它能够将短的 DNA 序列片段 (reads) 比对到人类基因组或其他较大的参考基因组上。它有两个版本：Bowtie 和 Bowtie2。Bowtie2 是 Bowtie 的升级版，能够比对更长的 reads，支持局部比对和剪接性比对。

Bowtie 的使用也需要先对参考基因组建立索引，然后再进行比对。比对的结果是一个 SAM 或 BAM 格式的文件，可以用 samtools 进行后续的处理。

Bowtie 和 BWA 都是常用的短序列比对工具，它们有一些相似之处，也有一些不同之处。

相似之处：

- 都是基于 Burrows-Wheeler Transform (BWT) 的算法，利用后缀数组和 FM-index 进行比对。- BWT 是 Burrows-Wheeler Transform 的缩写，它是一种数据转换算法，可以把一个文本转换成一个相似的文本，使得相同的字符更容易聚集在一起，从而方便后续的压缩。BWT 的原理是对文本的所有循环移位进行字典序排序，然后取最后一列作为转换后的文本。BWT 是可逆的，也就是说可以从转换后的文本恢复原文本。

- 都能够比对单端或双端的 reads，支持多种格式的输入和输出。
- 都能够处理比对错误和插入缺失，但是对于较长的插入缺失，效果不佳。

不同之处：

- BWA 有两种模式：BWA-MEM 和 BWA-ALN，前者适用于较长的 reads (> 70 bp)，后者适用于较短的 reads (< 70 bp)。Bowtie 有两个版本：Bowtie 和 Bowtie2，前者只支持全局比对，后者支持局部比对和剪接性比对。

- BWA 比 Bowtie2 更准确，但是 Bowtie2 比 BWA 更快。BWA 比 Bowtie 更敏感，但是 Bowtie 比 BWA 更节省内存。

- BWA 和 Bowtie2 都能够比对较长的 reads，但是 BWA-MEM 对于较长的 reads 更优于 Bowtie2。Bowtie 只能够比对较短的 reads (< 50 bp)。

- BWA 和 Bowtie2 都能够处理比对错误和插入缺失，但是 BWA-MEM 对于较短的插入缺失更优于 Bowtie2。Bowtie 只能够处理比对错误，不能处理插入缺失。

2 实验步骤

3 实验结果

4 实验总结