

# 天津医科大学实验课教案首页

(共 3 页、第 1 页)

课程名称：生物信息学	实验名称：实验九 制作人类基因剪接位点的 GT-AG 序列标识	
教师姓名：伊现富	职称：讲师	教学日期：2018 年 6 月 15 日 13:30-16:30
授课对象：基础医学院 2015 级基础班（本）	实验人数：19	
实验类型（验证型、综合型、设计型、创新型）：验证型	实验分组：一人一机	
学时数：3	教材版本：生物信息学实验讲义（自编教材）	

## 实验目的与要求：

- 掌握剪接位点的 GT-AG 法则。
- 掌握序列标识的含义。
- 掌握 Galaxy 的基本使用方法。
- 掌握 WebLogo 的基本使用方法。

## 实验内容及学时分配：

- (10') GT-AG 法则：回顾真核生物基因剪接位点的 GT-AG 法则。
- (10') 序列标识：回顾序列标识的含义。
- (10') WebLogo 简介：简单介绍 WebLogo 的作用及使用方法。
- (120') 实验操作：制作人类基因组中 22 号染色体上基因剪接位点的 GT-AG 序列标识。

## 主要仪器和实验材料：

- 实验材料：人类基因组（hg19）中 22 号染色体（chr22）上基因的剪接位点。
- 主要仪器：联网的计算机。
- 分析工具：Galaxy，WebLogo。

## 实验重点、难点及解决策略：

- 难点：序列标识的含义；解决策略：通过实例进行讲解。
- 重点：Galaxy 和 WebLogo 的使用；解决策略：根据资料进行学习，通过练习熟练掌握。

## 思考题：

- 真核生物基因剪接位点的 GT-AG 法则指的是什么？
- 序列标识的含义是什么？
- 如何制作序列标识？

## 参考资料：

- Galaxy
- WebLogo

主任签字：

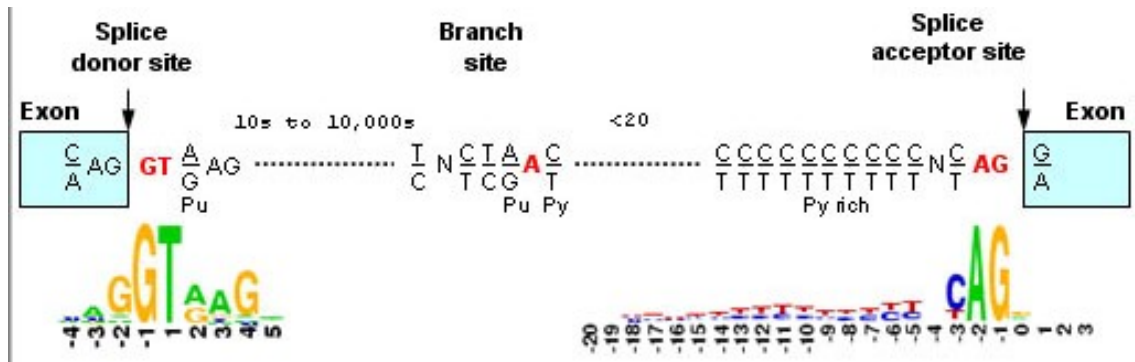
年 月 日

教务处制

## 一、GT-AG 法则 (10 分钟)

大多数真核基因的剪接位点都遵循“GT-AG 法则”：5'-GT...AG-3'。(注意：剪接位点的供体位点和受体位点都在内含子上！)

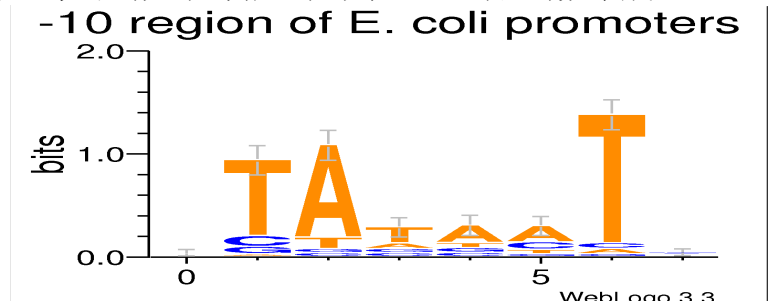
- 内含子 DNA 序列 5' 端起始的两个核苷酸是供体位点的 GT
- 内含子 DNA 序列 3' 端最后的两个核苷酸是受体位点的 AG
- GT 和 AG 这两个碱基序列具有高度保守性和广泛存在性
- A-G-[cut]-G-U-R-A-G-U (donor site) ...intron sequence ...Y-U-R-A-C (branch sequence 20-50 nucleotides upstream of acceptor site) ...Y-rich-N-C-A-G-[cut]-G (acceptor site)



## 二、序列标识 (10 分钟)

序列标识 (sequence logo) 是基于 DNA、RNA 和蛋白质的多序列比对信息，把多序列的保守性信息通过图形表示出来。它依次绘出基序中各个位置上出现的碱基，每个位置上所有碱基的累积可反映出该位置上碱基的一致性，每个碱基字母的大小与碱基在该位置上出现的频率成正比。

- 数据：多序列比对信息
- 横轴：序列坐标位置
- 纵轴：比特，计量单位
- 总高度：保守性
- 相对高度：相对频率
- 制作工具：WebLogo



## 三、WebLogo 简介 (10 分钟)

WebLogo is a web based application designed to make the generation of sequence logos easy and painless. WebLogo has featured in over 2000 scientific publications.

A sequence logo is a graphical representation of an amino acid or nucleic acid multiple sequence alignment developed by Tom Schneider and Mike Stephens. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

### WebLogo 3

#### Introduction

WebLogo is a web based application designed to make the generation of sequence logos easy and painless. WebLogo has featured in over 2000 scientific publications.

A sequence logo is a graphical representation of an amino acid or nucleic acid multiple sequence alignment developed by Tom Schneider and Mike Stephens. Each logo consists of stacks of symbols, one stack for each position in the sequence. The overall height of the stack indicates the sequence conservation at that position, while the height of symbols within the stack indicates the relative frequency of each amino or nucleic acid at that position. In general, a sequence logo provides a richer and more precise description of, for example, a binding site, than would a consensus sequence.

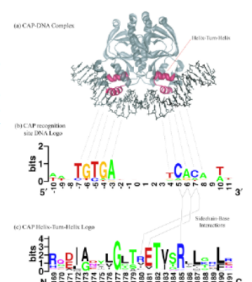
- [Create your own logos](#)
- [View example sequence logos and input data.](#)
- [Read the release notes for latest changes and updates.](#)
- [Read the User's Manual](#)
- [Download WebLogo source code](#)
- [WebLogo discussion group](#)

#### References

Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Research*. 14:1188-1190, (2004) [Full Text]

Schneider TD, Stephens RM. 1990. Sequence Logos: A New Way to Display Consensus Sequences. *Nucleic Acids Res.* 18:6097-6100

[about](#) · [create](#) · [examples](#) · [manual](#)



## 四、实验操作 (120 分钟)

制作人类基因组 (hg19) 中 22 号染色体 (chr22) 上基因剪接位点的 GT-AG 序列标识。

- 获取数据 (选择正确的格式, 把结果导出到 Galaxy 中)
  - 内含子数据: Get Data, UCSC Main, hg19, chr22, RefSeq Genes, Introns, BED 格式
- 提取剪接位点的信息 (比较并理解供体位点和受体位点的参数设置)
  - 32bp = 上游 15bp + 供体/受体位点 2bp + 下游 15bp
  - 供体位点的信息: Operate on Genomic Intervals, Get flanks, BED 格式
  - 受体位点的信息: Operate on Genomic Intervals, Get flanks, BED 格式
- 获取剪接位点的序列
  - 供体位点的序列: 基于供体位点的信息, Fetch Sequences, Extract Genomic DNA, FASTA 格式
  - 受体位点的序列: 基于受体位点的信息, Fetch Sequences, Extract Genomic DNA, FASTA 格式
- 多序列比对
  - 提取的序列已经是根据坐标比对好的序列, 没有必要再单独进行多序列比对了
  - 尝试进行多序列比对 (Multiple Alignments, ClustalW, FASTA 格式), 比较比对前后的结果 (如果出错, 请查找出错原因, 并尝试进行修正)
- 制作序列标识 (比较集成到 Galaxy 中的 WebLogo 和网页版的 WebLogo 的具体参数)
  - 集成到 Galaxy 中的 WebLogo (调整标题、图片格式等参数)
    - 供体位点的序列标识: 基于供体位点的序列, Motif Tools, Sequence Logo
    - 受体位点的序列标识: 基于受体位点的序列, Motif Tools, Sequence Logo
  - 网页版的 WebLogo (调整图片格式、First position number、标题等参数)
    - 下载剪接位点的序列
    - 供体位点的序列标识: 上传供体位点的序列
    - 受体位点的序列标识: 上传受体位点的序列
- 尝试制作同一物种或其他真核物种的不同染色体或全基因组基因的剪接位点的序列标识 (提示: 使用工作流简化重复操作)

