# 昆虫比较基因组学及化学感受受体基因家族的鉴定分析

[昆虫比较基因组学及化学感受受体基因家族的鉴定分析](#header-n0)  
 [1. 测序技术发展概述](#header-n21)  
 [1.1 第一代测序技术](#header-n24)  
 [Sanger](#header-n43)  
 [1.2 第二代测序技术](#header-n29)  
 [Illumina](#header-n61)  
 [1.3 第三代测序技术](#header-n33)  
 [PacBio & Nanopore](#header-n87)  
 [2. 昆虫基因组研究进展](#header-n125)  
 [3. 昆虫基因组测序](#header-n105)  
 [3.1 基因组大小评估](#header-n114)  
 [3.2 测序策略的选择](#header-n180)  
 [3.3 基因组DNA的提取及质控](#header-n134)  
 [4. 昆虫基因组组装](#header-n54)  
 [4.1 基因组组装软件及算法](#header-n70)  
 [4.2 影响基因组组装的因素](#header-n98)  
 [4.3 基因组组装的质量评估](#header-n100)  
 [5. 昆虫基因组注释](#header-n64)  
 [5.1 重复序列注释](#header-n161)  
 [5.2 非编码RNA注释](#header-n162)  
 [5.3 蛋白编码基因注释](#header-n418)  
 [6. 比较基因组学](#header-n66)  
 [6.1 基因家族聚类](#header-n200)  
 [6.2 物种进化分析](#header-n202)  
 [6.3 物种分歧时间估算](#header-n204)  
 [6.4 基因家族扩增和收缩分析](#header-n206)  
 [6.5 基因组共线性分析](#header-n76)  
 [7. 化学感受受体基因家族的鉴定](#header-n79)  
 [8. 未来和展望](#header-n464)

## 1. 测序技术发展概述

### 1.1 第一代测序技术

#### Sanger

### 1.2 第二代测序技术

#### Illumina

### 1.3 第三代测序技术

#### PacBio & Nanopore

## 2. 昆虫基因组研究进展

主要介绍昆虫基因组测序的数量，种类以及未来的研究方向。

## 3. 昆虫基因组测序

### 3.1 基因组大小评估

K-mer分析以及流式细胞分析

### 3.2 测序策略的选择

主要从测序物种基因组大小，基因组的杂合度，重复度以及体型大小展开讨论。

### 3.3 基因组DNA的提取及质控

包括测序材料的选择，DNA的提取量以及测序质量的控制。

## 4. 昆虫基因组组装

### 4.1 基因组组装软件及算法

Overlap-Layout-Consensus & De Bruijn graph

不同算法对应的不同软件

### 4.2 影响基因组组装的因素

重复序列的比例以及基因组的杂合度

### 4.3 基因组组装的质量评估

N50大小，组装的完整性和准确性

## 5. 昆虫基因组注释

### 5.1 重复序列注释

基于同源比对和从头预测的方法

### 5.2 非编码RNA注释

软件及数据库介绍

### 5.3 蛋白编码基因注释

结构注释：基于同源比对、从头预测以及转录组测序的方法

功能注释：与公共数据库同源比对，获取基因的功能信息、蛋白保守序列、结构域信息、生物学功能与过程信息及参与的生物通路信息等

## 6. 比较基因组学

### 6.1 基因家族聚类

使用 OrthoMCL/OrthoFinder 软件来进行基因家族聚类 。

### 6.2 物种进化分析

将所选物种的所有的单拷贝基因进行多序列比对，然后把所有的比对结果合并在一起，用于构建系统发育树

### 6.3 物种分歧时间估算

用单拷贝基因家族，使用 PAML 软件包中的 mcmctree 进行分歧时间估计

### 6.4 基因家族扩增和收缩分析

根据基因家族聚类结果和物种间的系统发育关系，使用 CAFE 进行基因家族扩张和收缩分析

### 6.5 基因组共线性分析

使用BLASTP 来检测与近缘物种间的直系同源基因，选择最优的比对结果，使用 MCscanX 识别同源基因区块，用于circos绘制物种间共线性图

## 7. 化学感受受体基因家族的鉴定

* 用tblastn将测序物种或近邻物种已鉴定的相关化学感受蛋白基因作为query，去搜索组装好的基因组；再用blastp将注释好的基因结构模型与NR数据库进行比对；最后用HMMER3将鉴定好的基因家族的基因与Pfam数据库进行比对获取基因的保守结构域。
* 将基因定位到染色体上，获得基因在染色体上的分布图谱。
* 与近邻物种的基因家族相关基因进行进化分析。

## 8. 未来和展望

介绍昆虫基因组测序的未来发展方向以及多组学数据的整合