**题目：寻找老鼠的基因组内源性逆转录病毒基因**

**摘要**

当逆转录病毒在种系DNA中形成时，它就变成了内生。至少有8%的人类基因组已经被发现逆转录病毒的来源。内源性逆转录病毒序列已在许多国家发现物种。在我们的研究中，我们搜索EST基因组的小鼠子集从已知的小鼠逆转录病毒蛋白质序列数据库。我们使用序列比对搜索结果与系统发生树和自组织定位来发现序列之间的模式和家族关系。

**方法**

在小鼠基因组中寻找MERVs

●从小鼠逆转录病毒基因组开始

●寻找三种逆转录病毒蛋白的保守区域(env, gag, pol)

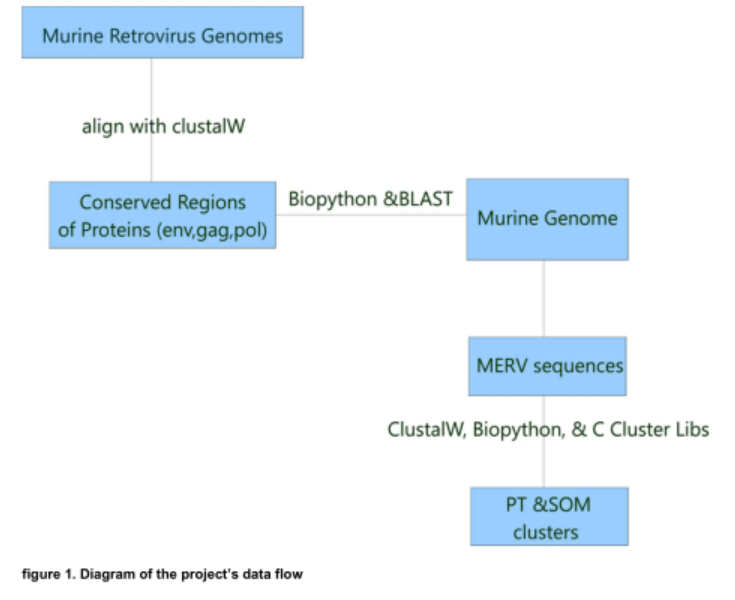
●使用保守区域作为探针“搜索项”，在其中搜索匹配项小鼠基因组，采用BLAST。

搜索所有基因组:

●采用小鼠逆转录病毒pol基因保守区作为探针

●利用该探针对整个NCBI基因数据库进行pcr扩增

Conserved sequences from murine retroviral genes:



小鼠逆转录病毒基因的保守序列:

聚类分析图与系统发育树

结论

我们发现:

●小鼠基因组中的逆转录病毒序列

●其他物种(猪、牛、人等)的小鼠逆转录病毒序列

●SOM聚类符合系统发育树聚类

实验理论

测定整个基因组，是找不到某个蛋白基因的。

对于某个功能蛋白基因，想要知道他的序列，首先要通过pcr扩增，得到DNA片段，但是如何设计引物呢？因为并不知道序列，所以只能在NCBI中通过该种相似功能蛋白的比对找到在该种蛋白基因序列中存在的保守序列（物种在长期进化过程中基本保持不变的序列），利用保守序列设计引物进行双端扩增，遇到终止子则停止扩增。这样就获得了蛋白基因序列。

本项目的思路是

1先在整个基因库中找到鼠类的逆转录病毒基因，进行序列比对，找到三种蛋白的逆转录病毒基因的保守区域。

2利用其中的保守序列与实验室测序的小鼠基因组进行比对，找到小鼠中的保守序列。

3选取其中的pol蛋白保守序列作为探针，在整个基因库进行pcr扩增（相当于寻找库中所有具备该小鼠逆转录病毒保守序列的基因），发现在其他物种中也存在该小鼠的逆转录病毒基因。

4 采用聚类分析，查看各种基因发育树关系。

Fasta文件是序列信息文件

通过clustalx序列两两比对生成aln文件，是比对结果文件。