**湖南师范大学生命科学学院学院**

**《生物信息学》课程期末上机试题**

(上机报告word文档请按下列方式命名：

专业+学号+姓名，如基地一201630151201李甜甜

（请务必写准确，老师下载后方便分班）

考试结束前将报告发到QQ邮箱：3432233705@qq.com)

**第一部分 生物信息数据库的检索**

1. 请在NCBI的核酸数据库中搜索斑马鱼的一段测序数据（Accession number详见数据使用库），并将全序列的FASTA文件下载到电脑。

**记录方式：对FASTA文件截图显示。**

**答：略**

1. 请通过NCBI的数据库检索工具检索此序列（同第一题），列出相应的查询结果。请截图参数设置界面、比对结果的图形化显示结果、比对结果列表（一部分即可）。并选取一组比对结果截图，指出该比对结果的得分和一致度。

**参数设置：**

Database：/nr核酸数据库

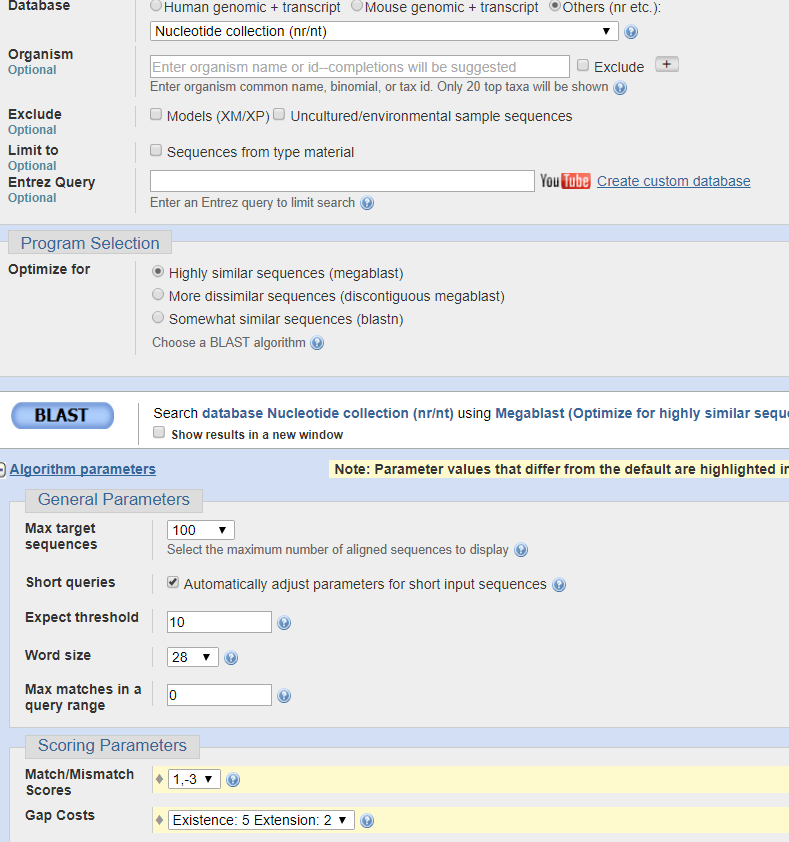
Program Selection：Highly similar sequences (megablast)

Algorithm parameters

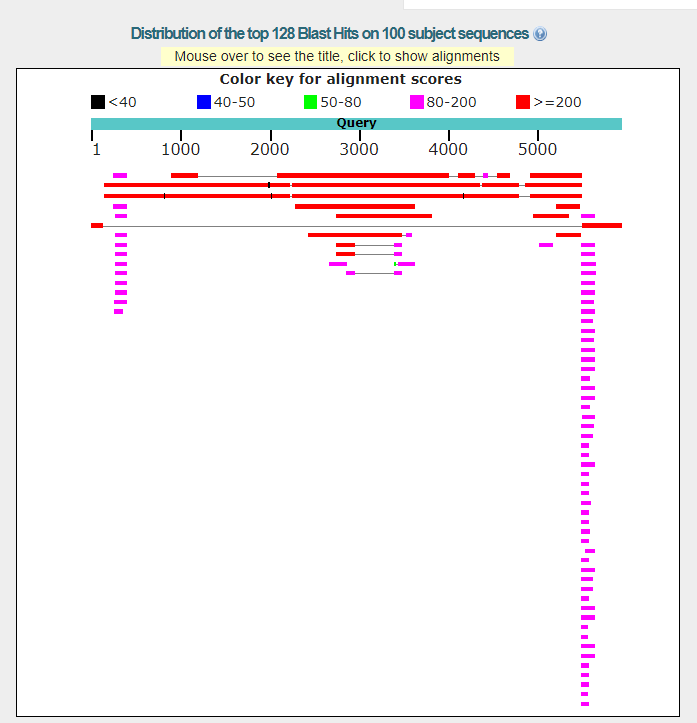
Word size:默认

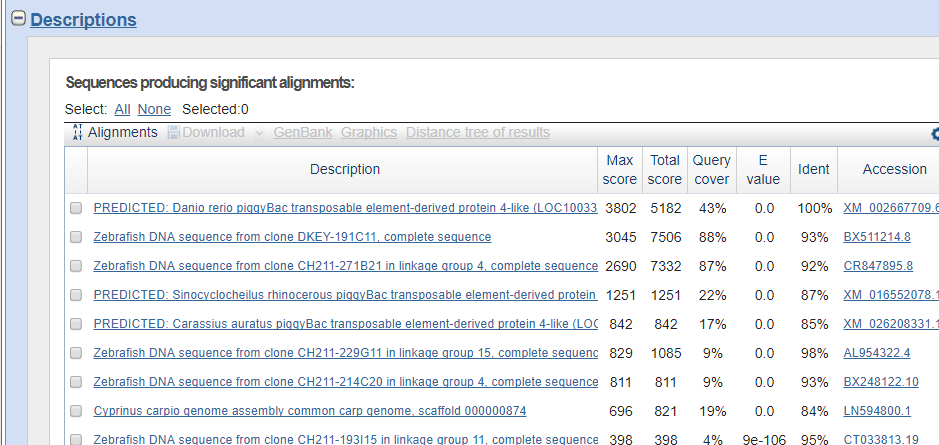
Match/Mismatch Scores:1,-3

Gap Costs: Existence:5， Extension:2

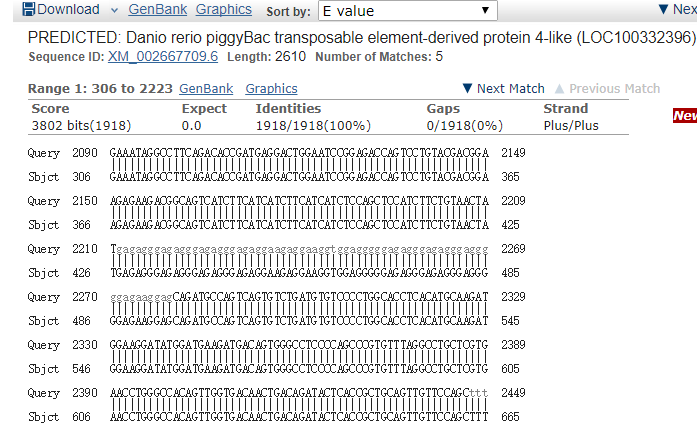
**答：参数设置**

**图形化结果：**



**对比结果列表**

**一组对比结果**



**得分: max score 3802;total score 5182 一致度:100%**

**第二部分 核酸序列特征分析**

1. 请使用工具对斑马鱼的核酸序列进行限制性内切酶酶切位点的分析（Accession number同第一题）。并观察该基因序列内是否被Agel和Hind III酶剪切（默认参数）。

记录方式：文字说明[Agel](http://nc2.neb.com/NEBcutter2/enz.php?name=919acf83-XM_005374233.2&enzname=AccI)和Hind III酶是否出现在序列中，并指出具体位点。如果有出现，请指出出现的具体位点和次数。

**答：Agel酶存在其剪切位点560共1个**

**Hind III酶存在其剪切位点1984, 2932, 3109, 4811共4个。**

1. 请使用GENSCAN在线工具分析对斑马鱼的一段核酸序列进行分析（Accession number与试题1相同），该基因结构由多少个外显子组成，编码多长的氨基酸序列？

参数设置：根据实际序列修改物种；

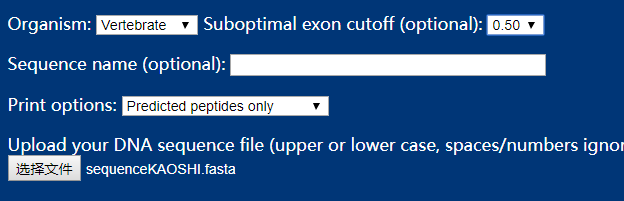
Suboptimal exon cutoff (optional):0.5。

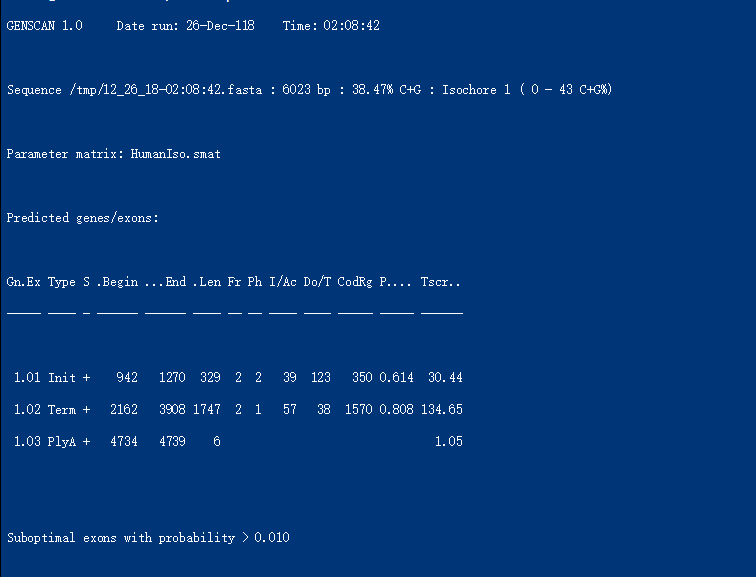
**记录方式：截图记录参数设置界面；**

**截图记录预测的最佳结果；**

**文字记录预测的外显子个数，编码氨基酸序列长度。**

**记录预测的氨基酸序列。**

**答：参数设置**

**预测的最佳结果**

**外显子个数:3 氨基酸序列长度691**

**氨基酸序列：**

MCVYTYYTCVNVLVTRAPSSVRVVCIERSETSLVLFTDRSPFRVFPTMDRQVQRSFCGVQ

SVLEEIMRSSDEEQQRDAETCFSDKESTATSVDSAQEDMFVDGLDPVLDRNRPSDTDEDW

NPETSPVRRKRRRQSSSSSSSSPAPSSVTMRGRGRGRGRGRWRGRGRGRGRRSRCQSVSD

VSPGTSHARWKDMDEDDSGPPQPVFRPAREPGPQLVTTDRYSPLQLFQLFFTPSVLQTIM

QHSNAYATKQRQGTDKSWKNISVEDLKSYLALVIFMGLLKVCALTDYWRRSDVYSVPLPT

RIMSCRKFLNIANALHLSDPVDDEANEGRKGTAAYDRLGKIKPLYDNIRDACRAFYHPGQ

NISIDERMVASKARSGLKQYMKNKPTKWGYKLFVLADSQCGYTSEFFVYEGKRMSSPNGL

SYESVMALIDEKYLGTGYKLYVDNFYTSPKLFRDLLSKRIYACGTIRPNRIGFPKMATNK

LPQDAPRGSMRWIREDDLLFVEWKDTREVLMCSTFHKAYGGDIVERKSKDNDGQWSTVHV

PIPGAVLDYNRFMGGVDLSDALICYYKVLHKTRKWYRTFFYHFVDIAVVNAFILHQKLAR

AKKERPLTQKAFRETLVMELAGLESGASPIPTPSTICHNPKYISGDSTSGRRKCRVCHQK

TPVMCATCEVPLCFLPKRDCYNDWHRREGKY

1. 你对题4中GENSCAN的预测结果不是很确信，请利用NCBI的BLAST工具，对GENSCAN预测的蛋白序列进化检索核对，根据检索结果，你认为GENSCAN的预测结果是否与事实相符？

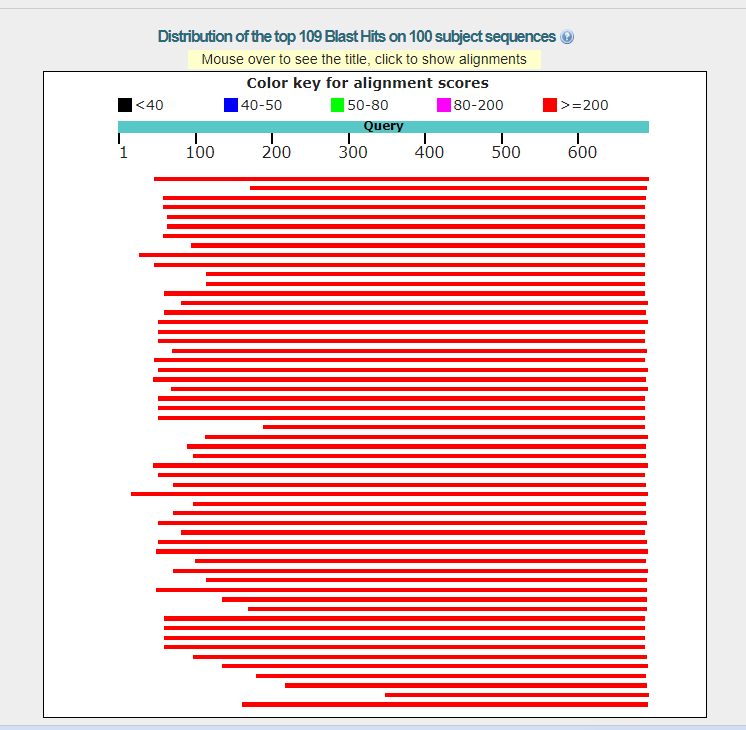
**Blast参数设置：**

Database：Non-redundant protein sequences(nr)

Program Selection： Quick BLASTP

其他参数默认。

**记录方式：请截图参比对结果的图形化显示结果。并文字叙述Genscan预测结果是否准确，简要说明原因。**

**答：图形化显示结果**

**结果准确；存在序列同一性达到100%，许多序列极为接近，且为斑马鱼的序列。**

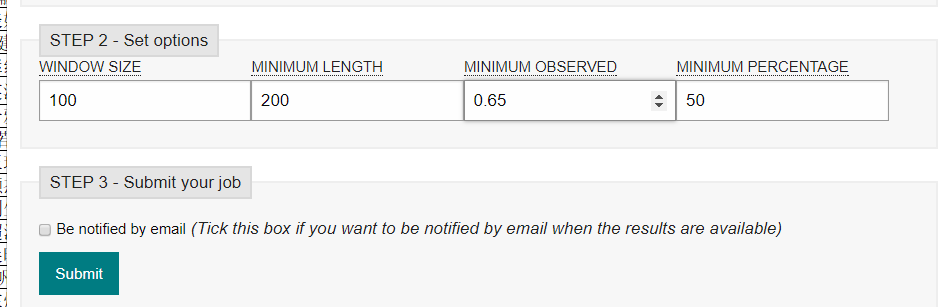
1. 请用[RepeatMasker](http://www.repeatmasker.org/)软件分析斑马鱼的这段核酸序列（Accession number同第一题），指出重复序列的类型，以及重复序列在查询序列中的位置信息。

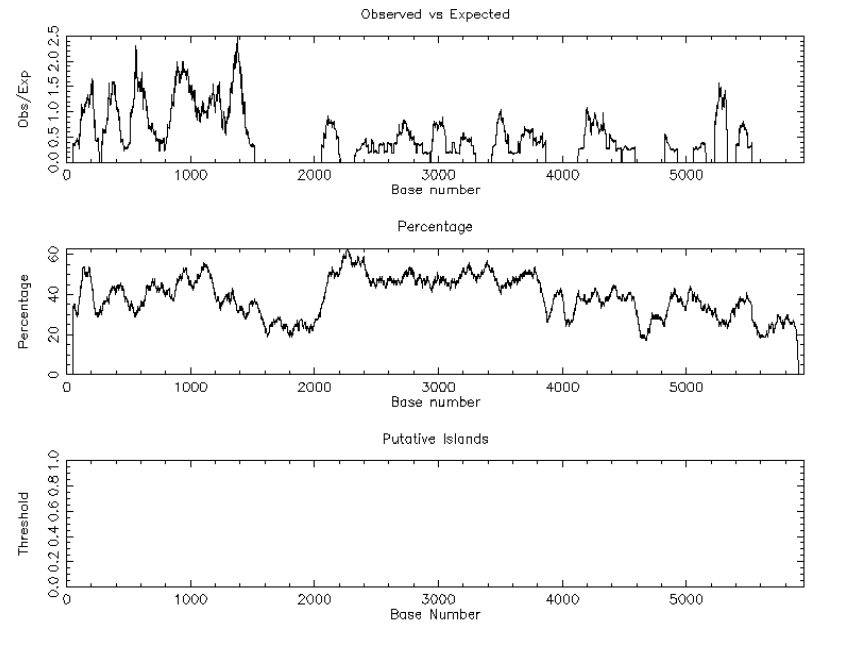
**答：存在5个重复序列，类型均为Simple\_repeat**

**位置信息分别为2164-2201；2211-2279 ；5559-5600；5601-5626 ；5933-5952。**

1. 请预测斑马鱼的这段核酸序列（Accession number同第一题）的CpG岛。

参数设置：窗口尺寸：100；最小CpG岛长度：200；observed/expected:0.65；最小CG含量：50%

**答：参数设置**

**预测结果**

**第四部分 进化树**

1. 将获取的对应序列整理为一个txt文档，按照序列格式要求修改为fasta格式（修改文件后缀名）；

利用mega软件，完成Clustalw比对后，利用ML进化方法制作系统发育树，用图片（TIFF格式）和进化树格式（Newick）保存，要求设置Bootstrap为500；

**提交结果**：

**1.序列比对整齐后的结果文件；**

**2．进化树结果（tiff图片和Newick格式）。**

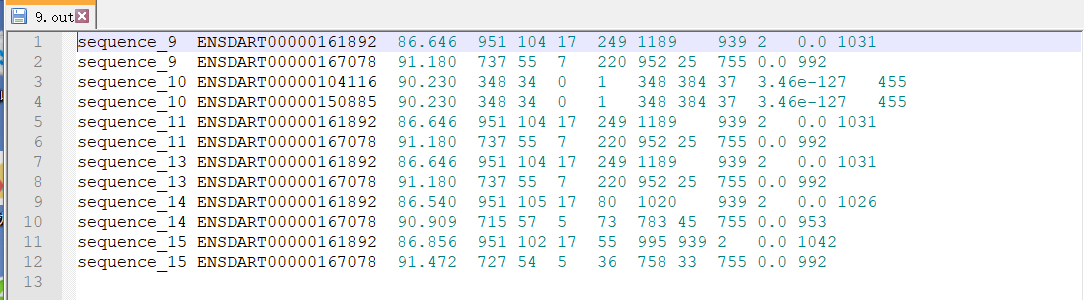
**第五部分 本地BLAST**

1. 将文件unknown sequence.txt文件中获得的序列（学号尾号两位数+9的数字的序列，比如学号尾号为01，获取第1条到第10条序列，学号尾号为02，获取第2条到第11条序列），利用本地blast软件比对数据库“zebrafish\_datebase.fasta”，完成对未知序列的本地blast注释；获取格式6的注释结果（使用参数 evalue = 1e-6; 最多比对2个注释结果），

**提交结果**：

**1.将最后的结果文件打开，将内容截图；**

**2.将建库和比对注释命令进行****截图。**

**答：1.** **结果文件**

2. **命令截图**

略