



电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

数据挖掘技术

Data Mining

期中报告

学生姓名 乔思尧

学 号 2011001010001

专 业 电子信息科学类

学 院 英才实验学校

指导教师 张会雄

2014年11月22日

数据挖掘技术期中 Project 报告

一、前言

数据生成和收集技术的进步促使商业和科研领域产生了海量数据集。数据仓库能够存储多重数据，如：企业销售和运作的详细情况，地球轨道卫星发送回地球的高分辨率图像和遥感数据，对越来越多的有机体进行的基因组实验产生的序列、结构和机能数据。收集和存储数据变得轻松简便，已经完全改变了人们对数据分析的态度，人们开始尽可能地收集各个时期和各种来源的数据。人们相信收集的数据肯定会有价值，或者当初收集它就有明确的目的，或者只能先收集起来再说。

传统数据分析技术在应对这些新型数据集提出的挑战时存在种种局限性，而数据挖掘技术突破了这些局限。数据挖掘并不是要取代其他分析领域，而是以它们为基础。尽管数据挖掘的某些主题（如关联分析）是其独有的，但是，还有许多主题（如聚类、分类和异常检测）则建立在其他领域长期工作的基础之上。事实上，数据挖掘研究者们主动利用已有技术对增强和拓展这个领域以及推动它的快速发展起到了促进作用。

该领域一直强调与其他领域的研究者合作，因而充满了活力。要迎接新类型数据分析的挑战，抛开理解数据的人和数据所处的领域而简单地使用数据分析技术是不可行的。通常，能否组建好多学科研究团队，已经成为数据挖掘项目（如创建新的独创性算法）成败的决定因素。正如历史上统计的许多进展都是由农业、工业、医疗卫生和商

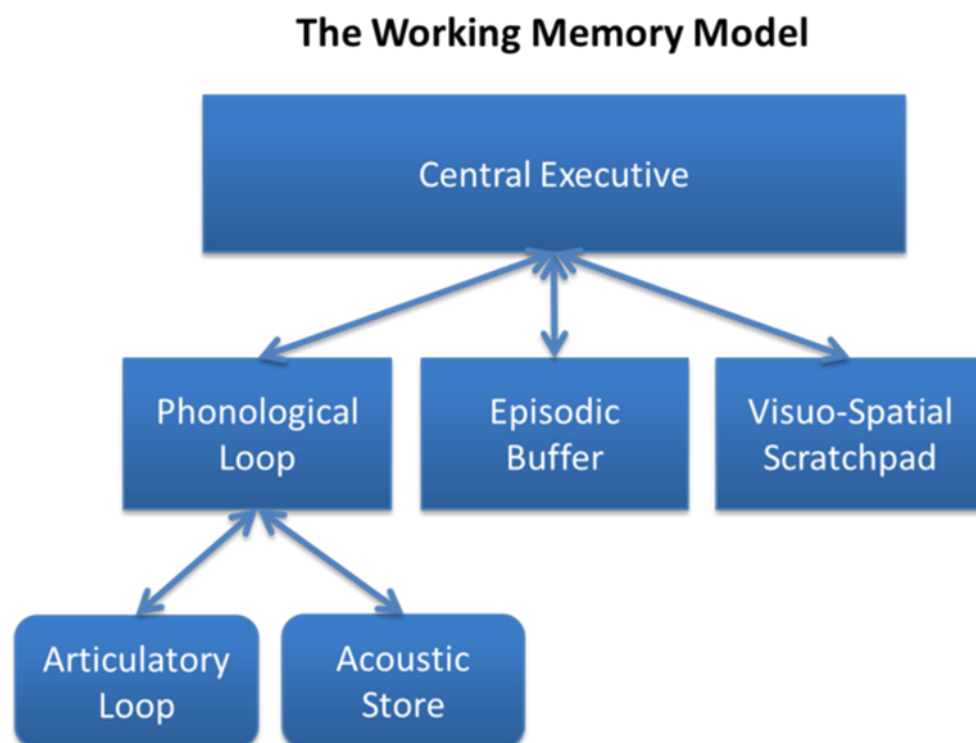
业需求推动的一样，今天，数据挖掘的许多进展也是正在被这些领域的需求所推动。

基于此，本篇报告将使用数据挖掘技术探寻急性心理应激与工作记忆之间的关系，以及说明如何通过数据挖掘技术获取草鱼一段未知基因的 DNA 序列。

二、急性心理应激与工作记忆

1、介绍

有很多证据表明慢性心理应激会通过损伤海马体对工作记忆产生不利影响，但急性心理应激对工作记忆的影响是有争议的。主流的观点认为急性心理应激也会损伤工作记忆，但也有小部分文献表明急性心理应激会提升工作记忆。

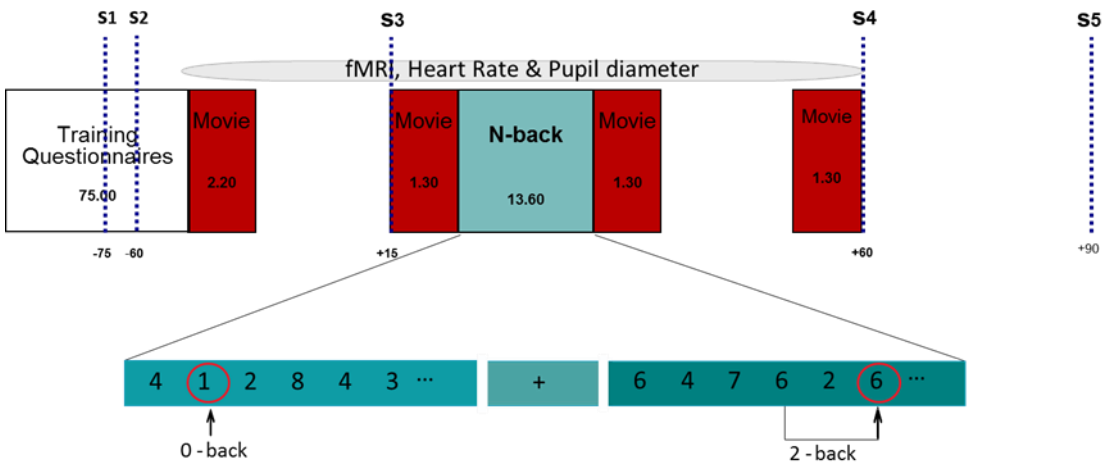


工作记忆是指在执行认知任务过程中，用于信息的暂时储存与加工的资源有限的系统，它是由额顶网络所支持的。急性心理应激可以

导致交感神经系统的迅速激活，同时伴随着前额叶皮质的去甲肾上腺素的大量释放。目前有学者研究表明，前额叶皮质是在工作记忆过程中起到重要作用的。并且，超适度的儿茶酚胺（如去甲肾上腺素）会抑制前额叶皮质，尤其是背外侧前额叶皮质的活性。因此我们猜想，急性心理应激对工作记忆可能会产生不利影响。此外，急性心理应激还会激活下丘脑-垂体-肾上腺轴，而这个激活作用会进一步损害背外侧前额叶皮质的作用。所以，我们预测急性心理应激可能会抑制背外侧前额叶皮质的活性从而对工作记忆产生不利影响。

为了解决这个问题，我们使用 fMRI 在急性心理应激情况下，进行 N-back 实验时神经活动是如何工作的。同时，为了监控交感神经系统和下丘脑-垂体-肾上腺轴是否被激活，我们还要收集唾液皮质醇浓度、心率以及心率变异性等数据。

2、实验流程



(1) 被试者

选择 27 名年龄大致相同、体质健康、惯用右手、视力正常的男性大学生进行实验。被试者要求没有神经疾病、精神疾病以及内分泌疾病，近段时间没有使用精神类和激素类药物，没有看恐怖电影或玩

恐怖游戏的习惯，没有严重的身体或心理创伤。在实验前两周，被试者将接受检测以确保激素分泌是正常的，并且填写知情同意书。然后，被试者将被随机分到实验组和对照组。

(2) 总体过程

实验从下午 2 点持续到 7 点，确保被试者的激素水平较为稳定。在进入 fMRI 扫描仪之前，被试者将进行一些工作记忆的训练并填写一些调查问卷，同时收集唾液皮质醇以备分析。然后被试者将被随机分到实验组 (n=14) 和对照组 (n=13)，但是他们自己并不知道自己在哪个组。在 fMRI 扫描仪中，被试者将看到 4 段电影片段，以此将被试者带入紧张(或中性)的情绪中。在第 2 个和第 3 个电影片段中间，被试者将进行工作记忆 N-back (0-back 和 2-back) 测试。

(3) 紧张诱导

实验组的被试者将看到 4 段恐怖电影片段，而对照组看到的是同样时间长度的无聊的电影片段。在观看过程中，被试者被要求全神贯注投入到电影片段的情节中，想象自己就是电影中的一员。

(4) N-back 测试

在第 2 段和第 3 段电影片段之间，被试者将进行 10 轮依次交替的 0-back 和 2-back 测试。在每一次测试中，被试者将看到 15 个随机的由个位数组成的数字序列，每一个数字将出现 400 毫秒，然后将是 1400 毫秒的间隔。在每一次测试开始之前，将有 2 秒的时间提示此次测试是 0-back 还是 2-back。在 0-back 的条件下，被试者被要求判断当前出现的数字是否是“1”；在 2-back 的条件下，被试者被

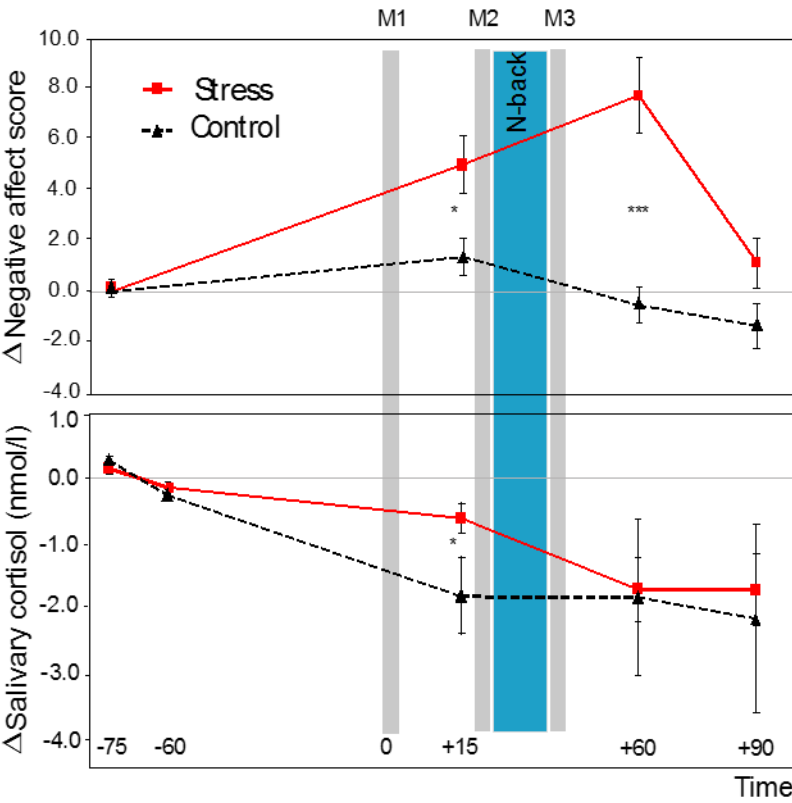
要求判断当前出现的数字是否与之前第二项的数字相同。在进入 fMRI 之前，被试者将进行 10 轮依次交替的 0-back 和 2-back 练习以减小变异性。最后 4 次练习的数据将被记录下来，作为不同被试者的衡量标准。

(5) 数据收集

主观情绪数据将通过正负性情绪测量表进行收集。为了监控下丘脑-垂体-肾上腺轴的活性变化，被试者的唾液皮质醇将被收集以备分析。同时，为了得到自主神经系统的响应变化，心率以及心率变异性也将被记录。在 N-back 实验过程中，被试者将进行 fMRI 脑部成像，以此分析脑部活动是否出现变化。

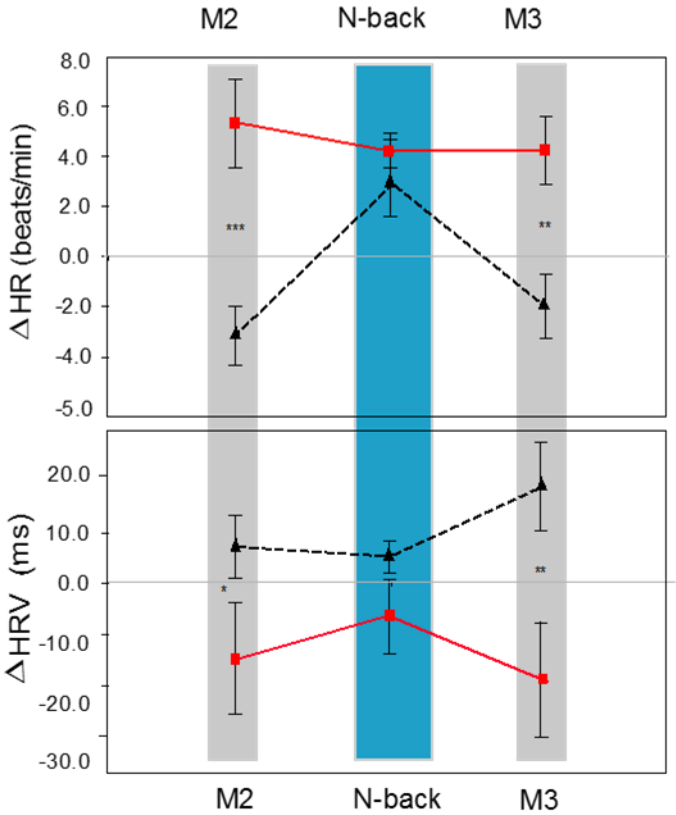
3、实验结果

(1) 数据分析



主观情绪数据经过组内和组间方差分析，结果显示随着时间的推进，两组之前有着极显著的差异，表明紧张情绪被成功地嵌入进实验组的被试者。

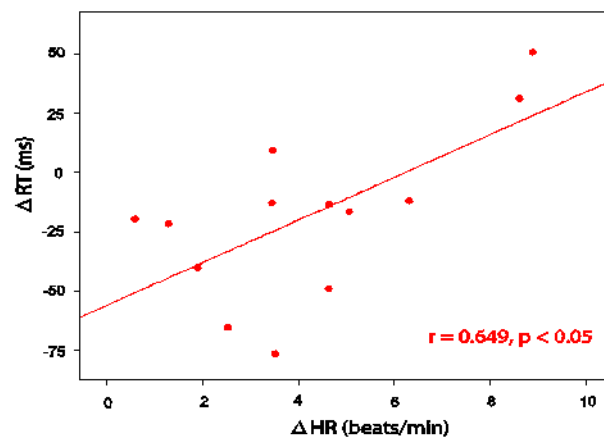
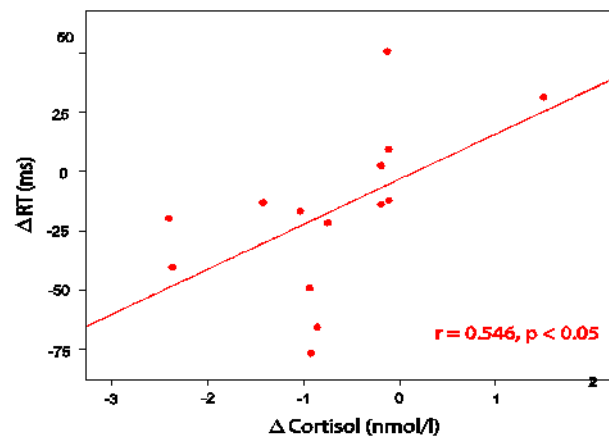
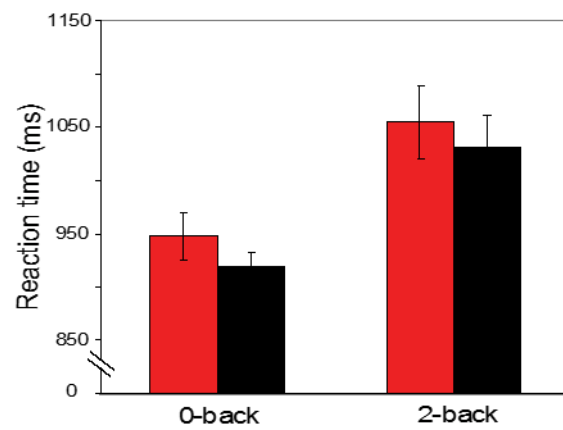
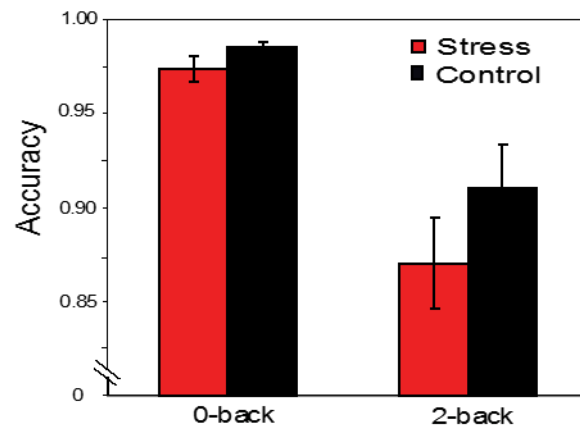
唾液皮质醇浓度经过方差分析，结果显示随着时间的推进，两个组都有所下降，这可能是正常生理规律导致的。而在 N-back 测试之前，实验组的数据明显高于对照组。



心率和心率变异性经过方差分析，结果显示随着时间的推进，两个组都出现了显著差异，其中心率的差异性在减小，而心率变异性在上升。

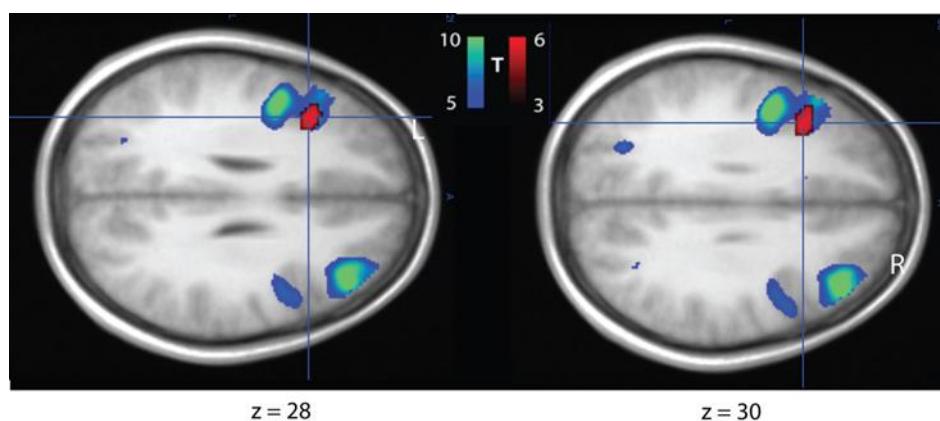
综合以上数据分析结果，我们可以确信在进行 N-back 测试时，紧张情绪以成功地嵌入进了实验组的被试者。

(2) N-back 测试



对测试的准确性和反应时间分别进行方差分析，我们发现实验组的准确性要低于对照组，而反应时间要高于对照组。进一步的，我们将唾液皮质醇浓度、心率以及反应时间分别做关联性分析，得出结论是随着唾液皮质醇浓度的上升或心率的上升，反应时间都会上升。因此，当被试者处于紧张的情绪时，反应能力就会下降。

（3）神经成像



我们扫描了以下区域：双边背外侧前额叶皮质、双边顶区内皮质、小脑等一些相关区域。我们发现在进行 N-back 测试时，实验组和对照组的背外侧前额叶皮质活性有着明显的差异，表明了急性心理应激情况下，背外侧前额叶皮质活性有着明显的降低。

4、实验结论

如同我们提出的猜想，在急性心理应激的情况下，背外侧前额叶皮质的活性将被抑制，原因可能是因为超适度的儿茶酚胺（如去甲肾上腺素）导致的。此外，唾液皮质醇浓度以及心率的变化显示出交感神经系统和下丘脑-垂体-肾上腺轴都被激活，从而进一步抑制背外侧前额叶皮质的活性。因此，在急性心理应激的情况下，工作记忆的确将会被抑制。

三、草鱼未知基因的分子克隆

1、实验任务

本次实验是在未知草鱼 STAT-1（信号转导及转录激活因子-1）基因序列的情况下，通过对其它鱼类 STAT-1 基因进行比对分析，推测出草鱼 STAT-1 基因的可能序列，并且通过分子克隆技术获取草鱼 STAT-1 基因全序列。

2、实验总体思路



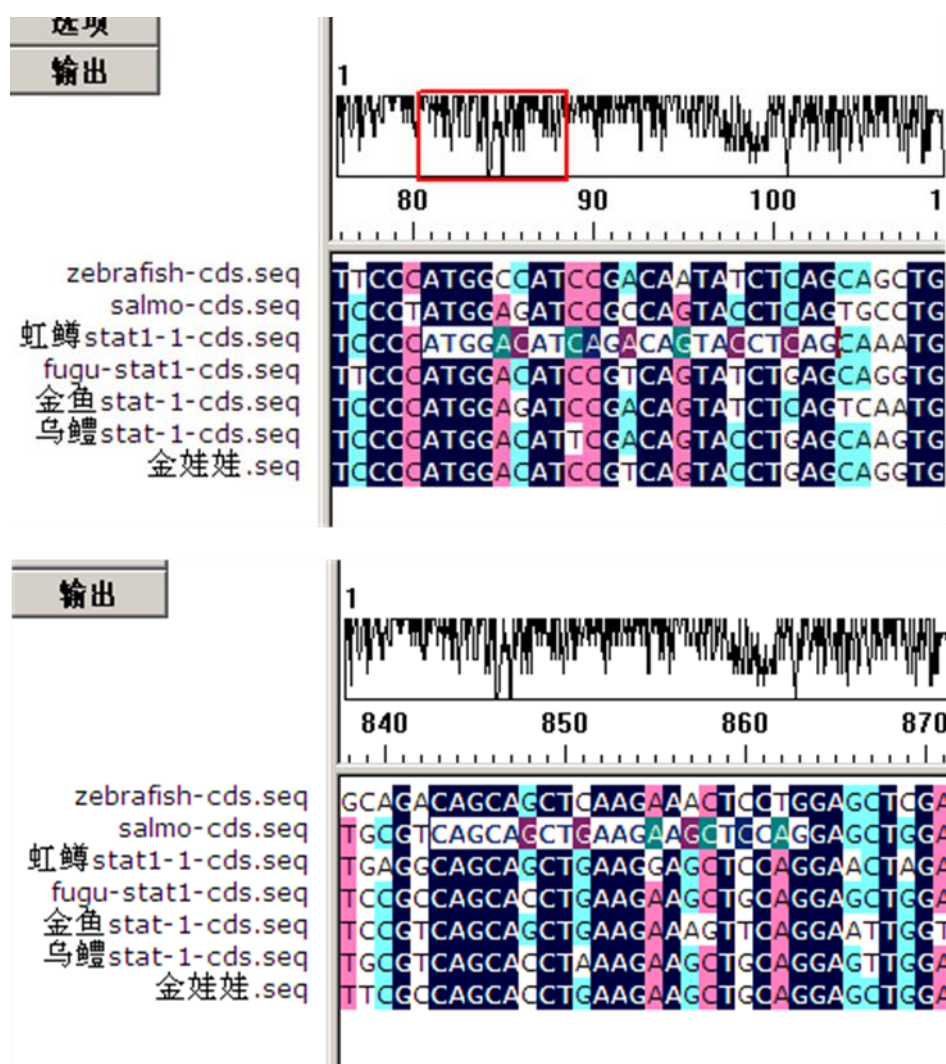
我们把这段 DNA 序列分为 3 个部分进行克隆。因为该 DNA 序列是未知的，我们不知道基因两端的序列，所以不可能一部克隆出完整的 DNA 序列。但是，根据 DNA 的保守性，我们可以通过比对其它鱼类的 STAT-1 基因，从中找出相似的部分设计出引物，从而通过该引物获取草鱼 STAT-1 基因的一段 DNA 序列，我们称这段序列为 Partial。得到结果后，我们再通过巢式 PCR 技术来分别克隆出 5' 端 3' 的 DNA 序列，经过拼接后得到该基因列。

3、实验完整步骤

（1）Partial 引物设计

由于所克隆的基因序列是未知的，所以我们只能通过 DNA 的保守性来设计我们需要的引物。通过查询，我们在 NCBI 网站上下载了斑马鱼、金鱼、日本河豚、虹鳟鱼、金娃娃等鱼类的 STAT-1 基因序列进行比对，选出其中相似程度较大的几个片段作为我们引物设计的备选片段。同时，我们设计的引物还要满足以下几个要求：

- ①引物的长度在 20-25bp 之间；
- ②TM 值在 66-70 之间；
- ③引物中 C/G 的个数在 40%-60%；
- ④引物最好以 C/G 开头和结尾；
- ⑤每个引物自身不要有互补性；
- ⑥引物之间的连续互补不超过 6bp 且互补处不能在 3' 段。



根据以上的要求，我们最终确定了 Partial 的两个引物：

F（前端引物）：ATGGA（C/G）ATCAG（A/T）CAGTA（C/T）CTCAG

R（后端引物）：CCTG（C/G）AG（T/C）TTCTTCAGCTGCTG

在设计出来之后，对引物之间的 bp 长度进行预测，为接下来的 PCR 实验做准备。我们的预测结果是 700bp。

(2) Partial 结果和分析

在设计出 Partial 引物以后，我们进行 PCR 实验，再经过跑胶、胶回收、转化和菌落 PCR，得到 Partial 的结果。拿到 Partial 的 DNA 序列以后，我们通过 NCBI 网站将其和其它鱼类的 STAT-1 基因进行比对，结果相似度较高，可以认为所得到的基因序列是我们所需要的 STAT-1 中的一段。

Description	Max score	Total score	Query cover	E value	Ident	Accession
Carassius auratus signal transducer and activator of transcription 1 (STAT1) mRNA, complete cds	1296	1296	100%	0.0	100%	AY242386.1
Danio rerio signal transducer and activator of transcription 1b (stat1b) mRNA, complete cds	792	792	100%	0.0	84%	FJ986224.1
PREDICTED: Danio rerio signal transducer and activator of transcription 1b (stat1b), transcript variant X1, mRNA	783	783	100%	0.0	84%	XM_005167878.1
Danio rerio signal transducer and activator of transcription 1b (stat1b), mRNA	783	783	100%	0.0	84%	NM_200091.2
Danio rerio signal transducer and activator of transcription 1b, mRNA (cDNA clone MGC:55503 IMAGE:2640987), complete cds	618	618	79%	3e-173	83%	BC044185.1
PREDICTED: Danio rerio novel protein similar to signal transduction and activation of transcription 1 (stat1) (LOC573407), misc_RNA	269	269	51%	2e-68	77%	XR_029517.4
Salmo salar Signal transducer and activator of transcription 1 (stat1), mRNA>gb BT048927.1 Salmo salar clone ssal-evf-519-071 Signal transducer an	244	244	99%	8e-61	68%	NM_001141285.1
Oncorhynchus mykiss clone omyk-evn-502-218 Signal transducer and activator of transcription 1 putative mRNA, complete cds	226	226	99%	2e-55	67%	BT073175.1
Oncorhynchus mykiss signal transducer/activator of transcription (stat1-1), mRNA>gb U60331.1 OMU60331 Oncorhynchus mykiss signal transducer/ac	223	223	98%	2e-54	68%	NM_001124707.1
Salmo salar STAT1a (STAT1a) mRNA, complete cds	208	208	98%	5e-50	68%	GQ325309.1
Salmo salar clone ssal-rgf-525-025 Signal transducer and activator of transcription 1-alpha/beta putative mRNA, complete cds	208	208	98%	5e-50	68%	BT045567.1
Salmo salar signal transducer and activator of transcription 1 alpha (LOC100136558), mRNA>gb EU016199.1 Salmo salar signal transducer and activ	208	208	98%	5e-50	68%	NM_001123654.1
Zebrafish DNA sequence from clone XX-85C7 in linkage group 19, complete sequence	203	1051	96%	2e-48	90%	AL592562.18
Zebrafish DNA sequence from clone CH211-196J1 in linkage group 9, complete sequence	203	1003	96%	2e-48	90%	BX649377.8
Zebrafish DNA sequence from clone BUSM1-8714 Contains part of a gene for a novel glutaminase, part of a gene similar to STAT1 (signal transducer an	203	1075	96%	2e-48	90%	AL672192.6
Zebrafish DNA sequence from clone BUSM1-199M19 in linkage group 7 Contains the 3' end of the Infa gene for tumor necrosis factor, alpha, the 3' end c	203	502	58%	2e-48	90%	AL596024.5
Zebrafish DNA sequence from clone IZAB-51F19 in linkage group 9 Contains the gene for a novel protein similar to vertebrate mitochondrial ribosomal r	203	738	84%	2e-48	90%	AL596027.11
PREDICTED: Oncotylagus cuniculus signal transducer and activator of transcription 1 (LOC100343893), mRNA	194	194	99%	1e-45	67%	XM_002712346.1
Mesocricetus auratus signal transducer and activator of transcription 1-like (LOC101827798), mRNA>dbj AB177397.1 Mesocricetus auratus STAT1 bet	176	176	93%	3e-40	67%	NM_001281685.1
Mesocricetus auratus signal transducer and activator of transcription 1 protein mRNA, complete cds	176	176	93%	3e-40	67%	DQ092343.1

(3) 5’ 与 3’ 基因序列

5’ 与 3’ 基因序列获取过程大致相同，这里以 5’ 为例来说一下我们所做的工作。

与之前克隆 Partial 段的 PCR 过程相似，不过我们需要 4 段引物来进行 PCR 实验。其中，5’ 段的两个引物是由生物公司提供的专门引物（5AAP 和 5AUAP），所以我们只需要从 Partial 的结果序列中选取符合引物要求的基因序列作为引物。引物的要求与 Partial 段中的

相同，在此不再赘述。

通过比对和分析，引物设计结果如下：

5N1: GCGTATATTATGCTGTAAGAGGAAG

5N2: GGATACATTACTTGCGACGGATTC

其中，5N1 是外侧的引物，5N2 是内测的引物。

(4) 5’ 与 3’ 结果和分析

与 Partial 段相似，我们进行巢式 PCR 实验（不是普通 PCR），再经过跑胶、胶回收、转化和菌落 PCR，得到 5’ 与 3’ 的结果。然后通过 NCBI 网站将其和其它鱼类的 STAT-1 基因进行比对，结果相似度较高，可以认为所得到的基因序列是我们所需要的 STAT-1 中的一段。

Description	Max score	Total score	Query cover	E value	Ident	Accession
Carassius auratus signal transducer and activator of transcription 1 (STAT1) mRNA, complete cds	360	360	96%	3e-96	100%	AY242386.1
Danio rerio signal transducer and activator of transcription 1b, mRNA (cDNA clone MGC-55503 IMAGE2640987), complete cds	251	251	94%	1e-63	89%	BC044185.1
PREDICTED: Danio rerio signal transducer and activator of transcription 1b (stat1b), transcript variant X1, mRNA	242	242	94%	7e-61	88%	XM_005167878.1
Danio rerio signal transducer and activator of transcription 1b (stat1b), mRNA	190	249	94%	4e-45	90%	NM_200091.2
Zebrafish DNA sequence from clone XX-85C7 in linkage group 19, complete sequence	190	258	94%	4e-45	90%	AL592562.18

Description	Max score	Total score	Query cover	E value	Ident	Accession
Carassius auratus signal transducer and activator of transcription 1 (STAT1) mRNA, complete cds	3714	3714	100%	0.0	99%	AY242386.1
PREDICTED: Danio rerio signal transducer and activator of transcription 1b (stat1b), transcript variant X1, mRNA	1669	1927	79%	0.0	87%	XM_005167878.1
Danio rerio signal transducer and activator of transcription 1b (stat1b), mRNA	1669	1927	79%	0.0	87%	NM_200091.2
Danio rerio signal transducer and activator of transcription 1b (stat1b) mRNA, complete cds	1663	1663	64%	0.0	87%	FJ986224.1
PREDICTED: Danio rerio novel protein similar to signal transduction and activation of transcription 1 (stat1) (LOC573407), misc. RNA	957	957	62%	0.0	75%	XR_029517.4
Salmo salar STAT1a (STAT1a) mRNA, complete cds	567	567	60%	2e-157	70%	GQ325309.1
Salmo salar clone ssal-rgf-525-025 Signal transducer and activator of transcription 1-alpha/beta putative mRNA, complete cds	567	567	60%	2e-157	70%	BT045567.1
Oncorhynchus mykiss signal transducer/activator of transcription (stat1-1), mRNA >gb U60331.1 JOMU60331 Oncorhynchus mykiss signal transducer/acti	567	567	60%	2e-157	70%	NM_001124707.1
Salmo salar signal transducer and activator of transcription 1 alpha (LOC100136558), mRNA >gb EU016199.1 Salmo salar signal transducer and activa	549	549	60%	4e-152	70%	NM_001123654.1
Oncorhynchus mykiss signal transducer/activator of transcription (stat1-2), mRNA >gb U60332.1 JOMU60332 Oncorhynchus mykiss signal transducer/acti	533	533	60%	3e-147	70%	NM_001124746.1
Paralichthys olivaceus signal transducer and activator of transcription 1 mRNA, complete cds	509	509	60%	4e-140	69%	EF491182.1
Siniperca chuatsi signal transducer and activator of transcription 1 mRNA, complete cds	489	489	60%	3e-134	69%	FJ629185.1
Epinephelus malabaricus signal transducer and activator of transcription 1 (STAT1) mRNA, complete cds	486	486	60%	4e-133	69%	KC992734.1
Homo sapiens mRNA for signal transducer and activator of transcription 1 isoform alpha variant, clone: FCC125G04	482	482	60%	5e-132	69%	AK225853.1
PREDICTED: Lepisosteus oculatus signal transducer and activator of transcription 1-like (LOC102685821), mRNA	475	475	53%	8e-130	70%	XM_006636664.1
PREDICTED: Pan troglodytes signal transducer and activator of transcription 1, 91kDa, transcript variant 2 (STAT1), mRNA	473	473	60%	3e-129	69%	XM_003309370.1
PREDICTED: Pan troglodytes signal transducer and activator of transcription 1, 91kDa, transcript variant 1 (STAT1), mRNA	473	473	60%	3e-129	69%	XM_003309369.1

4、实验结果

至此，我们已经成功获取了 Partial 段、5’ 端、3’ 端 3 段 DNA

序列，接着通过 3 段 DNA 序列之间重复的片段，将 3 段完整的 DNA 序列。我们再次通过 NCBI 网站将所得到的序列和其它鱼类的 STAT-1 基因序列进行比对，结果相似度较高，可以认为拿到了正确的序列。然后我们再使用高保真酶重复一次上述实验，最终获得草鱼 STAT-1 基因的完整序列。

5、附录

(1) Partial 序列

```
TCAATGGATCGAGAGTCATGACTGGGAATCCGTCGCAAGTAATGTATCCTTGGCC
ACCCTGCGCTTCCATGAACTCCTCAACCAACTGGATGAGCACTACAGCCGTCTCAACCT
GGGAAACAACCTTCCTCTTACAGCATAATATACGCAAAATTAAGCGGAACCTACAGGAGC
ACTTCCAGGAGGATCCTGTTACATGGCCATGATCATAGCTAATACTCTGAATGAAGAG
AGAAAGATACTGGAACTGCACTTAGTTTGCAAGGTAAAGGTGACTCATCACAAGGGAG
TTTCATGATGGAACAACAAAATGAGCTGGCCAATAAAATCAACAATTTGAAGACGAGCG
TGCAGGAAATAGAACAGGACATTCAGGTTTTGGAGGATGTACAAGATGAGTATGACTTC
AAGAGAAAGACCTTGCAGAGTCGAGTGGAAGCTGAAATGAATAGCCAGAAAACAAAAGA
GATCCAGCTGGAGGAAATGGCAATTCGACAAATGTTTATTGGGCTAGGCATGAAAAGAG
AGGTGGTGATAAGGGAAATGGCTAATGTACTGACTCTAGCAGAGCAGATCCAGTTCACA
CTCATTTTCAGATGAGTTACCCGAGTGGAAGAAGCGGCAACAGATGGCTTGCATCGGTGG
TCCGCCCAATGCATGTCTGGACCAGCTGCAGAGCTGGTTCACAGTGGTGGCAGAGTGTC
TGCAGCAGATCCGT
```

(2) 5' 端序列

```
GGGGGGGGACATGTATCTTGTCTGCTTATCCTGAGAAATACTGCTTTTGTGCCTT
```

GACCTAACTTCTATGTTGAAAGATGGCACTTTGGAACCAGCTGCAGCTGCTGGACTCTC
TGTATTTGGAGCAAGTGGACCAGCTTTACGATGAGGCTTTCCCATGGAGATCCGACAG
TATCTCAGTCAATGGATCGAGAGTCATGACTGG

(3) 3' 端序列

TTGCATCGGTGGCCCGCCCAATGCATGTCTGGACCAGCTGCAGAGCTGGTTCACA
GTGGTGGCAGAGTGTCTGCAGCAGATCCGTCAGCAGCTGAAGAAAGTTCAGGAATTGGT
GCAGAAGTTCACCTACAACAACGACCCCTCACTCTGGGCAAGAGCCGGCTGGATGAAC
AGGCCCTTTCCTCTTCAAAAACCTTATATTAAATTCTCTTGTGGTGGAGAGACAGCCG
TGTATGCCCACACATCCTCAGAGACCCTTGGTGATAAAGACAGGAGTTCAATTCACAGT
CAAGATCAGATCACTAGTCAAACCTCGCTGAATTGAACTGTCAACTCAAAGTAAAAGTCT
TTATTGACAAAGATTTAACAGAGAGTGACACACTAAAAGGATGTAGGAAATTCAACGTT
TTAGGAACAACGTCTAAAGTTATGAACTTAGAGGAATCCAATGGATGTCTAGCAGCTGA
ATTCGCCATTTGCAATTAAAGGAAATGAAATGTAACAACAGAACCAATGAGACTCCTC
TTATCATTTCTGAGGAGCTGCACTTACTTAGATTTGAACTCAACTCATTGAGCCAGAA
CTATGTGTTGATTTATCGATCATATCACTTCCTGTTGTGGTGATATCCACGTAAATCAG
TTACCTAGTGCTTGGGGCTCCATCTTATGGTACAACATGCTCTGCAGTGAACCACACAA
TCTGACGTTCTTCTTGAATCCGCCACCAGTGAAATGGGAGCAGCTTTCAAAGTCATAA
GCTGGCAGTTTTCTTCGGTCACGAAACGAGCGCTGAACTCTGAGCAGCTGAGAATGCTG
GCCAACAACTTCTTGGTCATGAAGCCCAAGGTGACCCTGAGGGTCTCATCTACTGGAA
TACATTTTGTAAGATGTCTTCTAATGAGAGGGGTGTACCATTCTGGCTGTGGATAGATG
GAATTCTGGACCTTATTAACCAACCTGCTCAACATCTGGAAAGATGGGCATATTATA
GGGTTTTTAAGTAAAGGGCAAGAGAAGGCTTTGTTGAGAGACAACTCCCAGGCACTTT

CCTCCTGCGCTTCAGTGAAACCTGTGCGAGAAGGAGGAATCACCATCACATGGGTGGAAT
ATTCAAAGAATGGTGAACCTAAGATGCACTCGGTGAAGCCTTACACTAAATCGGAACTG
GCGTCCATCTCCTTGCCTAACGTCATCCGGGACTACACCCTCACTGCTGCAGAGAAGGT
TCCAGTGAATCCCCTCATCTACCTTTACCCAGACATCCCCAAAAGATGAAGCTTTTAGTC
GCTACTATACCAGCGCACCTGATGATTCAGAAGAGATGGAGACAGACAACCCTATCAAG
CCATATATACCCAGACGCATGATTTCTGTGTCAGAAAACCCTGTTTCAAGACTCCAAAA
GTGTTAGAGCACATCGAGGCAACGCTGGTTTGAGGAACGTTTCTCCACCAATGCATGCT
GCATCTGACACTTAGATAACTATAATGAACGATGGAAACAAATAGTGATAACCCAAATGA
TTCATCTTTGCACAGGTCTTCTCTCTACTATATTAGTCATAAGAATGTACTACATTTAC
TCTCCACTTCTTGTGTGAGGAAGAGGAGCCATTTTCTTTTGTCAAATTACAGAATTTAT
AGTTAAAGCAGTGTAATCCTATTTCTTTTAGTTTGTTTGTTTTTTATAAAAAATGTGA
ATATATGGTCAAAAAACCTGTTGAGAATGCATTTTTTAGTCTGTGTACCAGTATTGCCA
TAATGGATACAAATACTCTGAAAATCTGTTTCTTTCAAGACATTTGTTTCGTGAACAGGA
CATCCTCCTCTATCTTGTGAAGGACCTATTTATTTTTGCACATCAAAATATCTTTACTG
TATATACCATTGCCTTAATTGCACAAACACACACACACTGATTTCAAGTCAGCCAGGTAC
AGTCCAACAGAAATAACAACAACCATCCTTCTCCAGTCAGAGATACTGCTTCAGATTC
GAGCCGGAATCATTTTGTTTCTATGTTTTGAACATTAGACAGTGTTTCATAATGTGAAT
TACTAATAAAGACTATTAAATGCTTTATAAGAAAAAAAAAAAAAAAAAAAA

(4) STAT-1 全序列

GGGGGGGACATGTATCTTGTCTGCTTATCCTGAGAAATACTGCTTTTGTGCCTT
GACCTAACTTCTATGTTGAAAGATGGCACTTTGGAACCAGCTGCAGCTGCTGGACTCTC
TGTATTTGGAGCAAGTGGACCAGCTTTACGATGAGGCTTTCCCATGGAGATCCGACAG

TATCTCAGTCAATGGATCGAGAGTCATGACTGGGAATCCGTCGCAAGTAATGTATCCTT
GGCCACCCTGCGCTTCCATGAACTCCTCAACCAACTGGATGAGCACTACAGCCGTCTCA
ACCTGGGAAACAACTTCCTCTTACAGCATAATATACGCAAAATTAAGCGGAACCTACAG
GAGCACTTCCAGGAGGATCCTGTTACATGGCCATGATCATAGCTAATACTCTGAATGA
AGAGAGAAAGATACTGGAACTGCACTTAGTTTGCAAGGTAAAGGTGACTCATCACAAG
GGAGTTTCATGATGGAACAACAAAATGAGCTGGCCAATAAAATCAACAATTTGAAGACG
AGCGTGCAGGAAATAGAACAGGACATTCAGGTTTTGGAGGATGTACAAGATGAGTATGA
CTTCAAGAGAAAGACCTTGCAGAGTCGAGTGGAAGCTGAAATGAATAGCCAGAAAACAA
AAGAGATCCAGCTGGAGGAAATGGCAATTCGACAAATGTTTATTGGGCTAGGCATGAAA
AGAGAGGTGGTGATAAGGGAAATGGCTAATGTACTGACTCTAGCAGAGCAGATCCAGTT
CACACTCATTTCAGATGAGTTACCCGAGTGGAAGAAGCGGCAACAGATGGCTTGATCG
GTGGTCCGCCCAATGCATGTCTGGACCAGCTGCAGAGCTGGTTCACAGTGGTGGCAGAG
TGTCTGCAGCAGATCCGTCAGCAGCTGAAGAAAGTTCAGGAATTGGTGCAGAAGTTCAC
CTACAACAACGACCCCCTCACTCTGGGCAAGAGCCGGCTGGATGAACAGGCCCTTTAC
TCTTCAAAAACCTTATATTAAATTCTCTTGTGGTGGAGAGACAGCCGTGTATGCCCACA
CATCCTCAGAGACCCTTGGTGATAAAGACAGGAGTTCAATTCACAGTCAAGATCAGATC
ACTAGTCAAACCTCGCTGAATTGAACTGTCAACTCAAAGTAAAAGTCTTTATTGACAAAG
ATTTAACAGAGAGTGACACACTAAAAGGATGTAGGAAATTCAACGTTTTAGGAACAACG
TCTAAAGTTATGAACTTAGAGGAATCCAATGGATGTCTAGCAGCTGAATTTGCCATTT
GCAATTAAAGGAAATGAAATGTAACAACAGAACCAATGAGACTCCTCTTATCATTTCTG
AGGAGCTGCACTTACTTAGATTTGAACTCAACTCATTCAGCCAGAACTATGTGTTGAT
TTATCGATCATATCACTTCCTGTTGTGGTGATATCCACGTAAATCAGTTACCTAGTGCT

TGGGGCTCCATCTTATGGTACAACATGCTCTGCAGTGAACCACACAATCTGACGTTCTT
CTTGAATCCGCCACCAGTGAAATGGGAGCAGCTTTCAAAAGTCATAAGCTGGCAGTTTT
CTTCGGTCACGAAACGAGCGCTGAACTCTGAGCAGCTGAGAATGCTGGCCAACAACTT
CTTGGTCATGAAGCCCAAGGTGACCCTGAGGGTCTCATCTACTGGAATACATTTTGTA
GATGTCTTCTAATGAGAGGGGTGTACCATTCTGGCTGTGGATAGATGGAATTCTGGACC
TTATTA AAAACACCTGCTCAACATCTGGAAAGATGGGCATATTATAGGGTTTTTAAGT
AAAGGGCAAGAGAAGGCTTTGTTGAGAGACAACTCCCAGGCACTTTCCTCCTGCGCTT
CAGTGAAACCTGTGCGAGAAGGAGGAATCACCATCACATGGGTGGAATATTCAAAGAATG
GTGAACCTAAGATGCACTCGGTGAAGCCTTACACTAAATCGGAACTGGCGTCCATCTCC
TTGCCTAACGTCATCCGGGACTACACCCTCACTGCTGCAGAGAAGGTTCCAGTGAATCC
CCTCATCTACCTTTACCCAGACATCCCAAAGATGAAGCTTTTAGTCGCTACTATACCA
GCGCACCTGATGATTCAGAAGAGATGGAGACAGACAACCCTATCAAGCCATATATACCC
AGACGCATGATTTCTGTGTCAGAAAACCCTGTTTCAAGACTCCAAAAGTGTTAGAGCAC
ATCGAGGCAACGCTGGTTTGAGGAACGTTTCTCCACCAATGCATGCTGCATCTGACACT
TAGATAACTATAATGAACGATGGAAACAAATAGTGATACCCAAATGATTCATCTTTGCA
CAGGTCTTCTCTCTACTATATTAGTCATAAGAATGTACTACATTTACTCTCCACTTCTT
GTGTGAGGAAGAGGAGCCATTTTCTTTTGTCAAATTACAGAATTTATAGTTAAAGCAGT
GTAATCCTATTTCTTTTAGTTTGTTTGTATATAAAAAATGTGAATATATGGTCAA
AAAACCTGTTGAGAATGCATTTTTTAGTCTGTGTACCAGTATTGCCATAATGGATACAA
ATACTCTGAAAATCTGTTTCTTTCAAGACATTTGTTTCGTGAACAGGACATCCTCCTCTA
TCTTGTGAAGGACCTATTTATTTTGCACATCAAATATCTTTACTGTATATACCATTG
CCTTAATTGCACAAACACACACACACTGATTTCAAGTCAGCCAGGTACAGTCCAACAGAA

ATAACAACAACCATCCTTCTCCAGTCAGAGATACTGCTTCAGATTCGAGCCGGAATCA
TTTTGTTTCTATGTTTTGAACATTAGACAGTGTTTCATAATGTGAATTACTAATAAGA
CTATTAAATGCTTTATAAGAAAAAAAAAAAAAAAAA