**实验一·海洋生物功能基因生物信息学分析**

1. 实验目的：生物信息学作为一门新兴的交叉学科，有其独特的优势及发展空间。生物信息学是一个建立在对DNA和蛋白质序列比较基础上的学科，目的是发现进化关联，并由此进行功能比较。本实验旨在对海洋生物功能基因的基因及蛋白质等结构进行分析。
2. 实验原理：利用生物信息学分析软件及相关网站对海洋生物功能基因的序列的生物信息进行处理、分析和解释。
3. 实验材料：
4. 生物信息学分析软件及相关网站。
5. 目的基因检索与分析。
6. 主要实验方法：
7. **基因全长序列分析：**对于给定的DNA序列进行分子生物学分析。通过NCBI的Open Reading Frame Finder（<http://www.ncbi.nlm.nih.gov/gorf/gorf.html>）分析可知：给定的基因序列全长，及能编码氨基酸数量。
8. **蛋白理化性质预测与分析：**根据在线服务系统ExPASy中的ProtParam工具[(http://www.expasy.org/tools/protparam.html)](http://www.expasy.org/tools/protparam.html)和Proscale工具<(http://webexpasy.org/ cgi-bin/protscale/protscale.pl)>进一步对目的基因蛋白氨基酸序列的基本理化性质进行综合预测分析，结果比较可靠。预测的理化性质有：蛋白分子量(Molecular weight)、理论等电点(Theoretical pI)、氨基酸组成(Amino acid composition)、电荷分布(negatively charged residues, positively charged residues)、原子构成(Atomic composition )、消光系数( Extinction coefficients )、半衰期(Estimated half-life)、不稳定系数(Instability index)，脂肪系数(Aliphatic index)、总平均疏水性(Grand average of hydropathicity, GRAVY )等。
9. **磷酸化位点预测与分析：**磷酸化和去磷酸化是细胞内信号传导的重要方式，应用在线服务NetPhos 2.0 Serve[(http://www.cbs.dtu.dk/services/NetPhos/)](http://www.cbs.dtu.dk/services/NetPhos/)对目的基因的磷酸化位点进行预测与分析，对蛋白序列中的Ser、Thr和Tyr三种氨基酸残基可能成为的磷酸化位点作出预测。
10. **蛋白卷曲螺旋结构的预测与分析：**利用在线服务Coils分析工具<(http://embnet. vital-it.ch/software/COILS_form.html)>对目的基因蛋白序列形成卷曲螺旋的倾向性进行预测，以window=14，21和28为实验参数，按照几率>50%就可形成螺旋的规则，比较不同权重情况下的分析结果。
11. **蛋白质二级结构预测与分析：**蛋白质二级结构主要有α-螺旋、β-折叠、β-转角等几种形式，它们是构成蛋白质高级结构的基本要素。应用JPred[（http://www.compbio.dundee.ac.uk/jpred/）](http://www.compbio.dundee.ac.uk/jpred/)和PredictProtein [(www.predictprotein. org/)](http://(www.predictprotein. org/))预测服务器对目的基因蛋白进行二级结构预测。
12. **蛋白的跨膜结构预测与分析：**跨膜结构域的预测和分析，对正确认识和理解蛋白质的功能、结构、分类、方位及细胞中作用部位均有着重要的指示意义。应用在线服务TMHMM [(http://www.cbs.dtu.dk/services/TMHMM/)](http://www.cbs.dtu.dk/services/TMHMM/)和Tmpred [(http://www.ch.embnet.org/software/TMPRED\_form.html)](http://www.ch.embnet.org/software/TMPRED_form.html)对目的基因蛋白的跨膜结构进行预测及分析。红色表示跨膜区，蓝色即在膜内部，相反紫色细线表示在膜外的概率。
13. **信号肽的预测和分析：**预测和分析信号肽有助于蛋白质功能域的划分和蛋白质细胞定位。应用在线服务SignaIP 4.1 Server[(http://www.cbs.dtu.dk/services/SignalP/)](http://www.cbs.dtu.dk/services/SignalP/)对目的基因的信号肽进行预测与分析。
14. **亚细胞定位预测与分析：**亚细胞定位与蛋白质的功能存在着非常重要的联系，可以通过氨基酸组成进行亚细胞定位的预测。应用PSORT[(http://psort.nibb.ac.jp/)](http://psort.nibb.ac.jp/)软件对目的基因蛋白的亚细胞内定位进行预测。[这个网站需要翻墙才能登陆，可以试一下，不一定能打开。]
15. **三维结构的预测和分析：**蛋白质依赖于其三维结构的形状和关键功能域的性质来实行生物功能。利用Phyre在线工具[(http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)](http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index)对目的基因氨基酸序列进行蛋白质三维结构预测。
16. **保守结构域与功能域分析：**根据NCBI-CDS[(http://www.ncbi.nlm.nih.gov/cdd/)](http://www.ncbi.nlm.nih.gov/cdd/)和Prosite [(http://prosite.expasy.org/scanprosite/)](http://prosite.expasy.org/scanprosite/)对目的基因蛋白的保守结构域与功能域进行在线分析。
17. **同源蛋白质家族比较分析**：应用InterProScan程序搜索位于EBI的InterPro数据库[（http://www.ebi.ac.uk/interpro/search/sequence-search）](http://www.ebi.ac.uk/interpro/search/sequence-search)进行同源蛋白质分析比较。
18. **分子系统发育预测与分析：**
19. 首先对目的基因的氨基酸序列进行BLAST分析。应用NCBI的BLAST在线工具（<http://blast.ncbi.nlm.nih.gov/Blast.cgi>）进行在线BLAST分析，检索并选取若干与给定目的基因相似度较高的基因进行下载，尽量不要选带SP.（未知菌）的菌株基因，选择typical菌株的基因，并且选取时主要不要选择染色体基因。下载时选定Fasta格式，得到txt文件，打开该文件，并进行编辑，主要是菌株及基因名称及genbank编号等。
20. 打开BioEdit软件，并打开下载好的txt文件。选择Accessory Application，选择ClustalW Multiple alignment，及Run ClustalW，保存文件untitled.fas。
21. 打开MEGA软件，打开untitled.fas文件，确认进行分析。选择Nucleotide Sequences，并确认，不进行核苷酸序列的蛋白质编码。选择phylogeny，选择第二个Construct/Test Neighbor-Joining Tree即可。之后进行保存，并可用AI等软件进行修饰。
22. **基因功能的预测与分析**：应用在线软件InterPro <(http://www.ebi.ac.uk/ interpro/search/)>对目的蛋白参与的生物过程、分子功能进行预测与分析。

注意事项：所有在线分析网站及软件大部分需要Fasta格式，即为在给定的目的基因的前加上“>”及基因名称，否则无法进行分析。