

实 验 报 告

|  |  |  |
| --- | --- | --- |
| 课程名称 | ： | 生物信息学实验 |
| 课程号 | ： |  |
| 实验学期 | ： | 2024-2025学年2学期 |
| 学院 | ： | 农学院 |
| 班级 | ： | 智慧农业 |
| 学号 | ： |  |
| 姓名 | ： |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 实验项目 | 一 | 二 | 三 | 四 | 五 | 六 | 七 | 八 | 九 |
| 得分 |  |  |  |  |  |  |  |  |  |
| 指导教师 |  |  |  |  |  |  |  |  |  |
| 总成绩 | |  | | | | | | | |

注：请同学们实验前扫描二维码，学习《东北农业大学实验室安全知识手册》。



东北农业大学实验室安全知识手册二维码

实 验 报 告

|  |  |  |  |
| --- | --- | --- | --- |
| 实验项目名称：**实验一、常用生物信息数据库检索** | | | |
| 实验时间 | 2023-2024年  2学期第11-13周 | 同组人数 | 1 |
| 实验地点 | 成艺楼309 | 试验台号 | 无 |
| **一、实验目的：**  1、熟悉常用生物信息数据库的数据类型、数据格式及信息内容；  2、掌握常用生物信息数据库的检索方法。  **二、实验材料：**  可访问互联网的计算机。  **三、实验主要步骤与结果：**  **Ⅰ Nucleotide数据库的检索**  ① 打开浏览器并访问NCBI主页  URL: https://www.ncbi.nlm.nih.gov/  ② 进入NCBI Nucleotide数据库  在NCBI主页的搜索框中，点击左侧的下拉菜单，选择“Nucleotide”数据库。  输入关键词进行搜索  ③ 在搜索框中输入关键词：Oryza sativa drought resistance gene。  点击“Search”按钮进行搜索。  ④ 筛选搜索结果  在搜索结果页面中，浏览与抗旱性相关的基因序列。可以根据需要在页面右侧的“Filters”中使用筛选条件（如：基因种类、来源物种、序列长度等）以精确查找目标基因。  点击相关的基因链接，进入具体的基因信息页面。  ⑤ 查看并分析基因序列  在基因信息页面中，切换到“Sequence”选项卡查看该基因的完整序列。  选择“FASTA”格式查看基因序列，方便后续的生物信息学分析。  ⑥ 下载基因序列  在基因页面的右上角，点击“Send to”按钮。  在弹出的选项中选择“File”，并选择“FASTA”格式，然后点击“Create File”下载基因序列。  注意：在上述过程中描述中选择2张有代表性的过程截图和1张结果截图，注意只截取网页内容部分，其他部分不要截取，下同。  **Ⅱ NCBI中蛋白质结构查询**  ① 打开浏览器并访问NCBI主页  URL: https://www.ncbi.nlm.nih.gov/  ② 进入NCBI Protein数据库  在NCBI主页的搜索框中，点击左侧的下拉菜单，选择“Protein”数据库。  ③ 输入关键词进行搜索  在搜索框中输入关键词：Arabidopsis thaliana photosynthesis protein。  点击“Search”按钮进行搜索。  ④ 筛选搜索结果  在搜索结果页面浏览与光合作用相关的蛋白质条目。  ⑤ 查看蛋白质序列与结构信息  在蛋白质页面的“Sequence”部分查看蛋白质的氨基酸序列。  ⑥ 下载蛋白质序列或结构  下载蛋白质序列，在“Send to”菜单中选择“File”并选择“FASTA”格式，然后点击“Create File”进行下载。  如要下载蛋白质的三维结构，需要进入PDB页面，选择“Download Files”并下载“PDB Format”文件，方便后续进行结构分析。  **Ⅲ 表达谱数据检索**  ① 打开浏览器并访问NCBI GEO主页  URL: https://www.ncbi.nlm.nih.gov/geo/  ② 进入GEO数据库的GEO DataSets页面  点击页面中的“GEO DataSets”进入数据集检索页面，或直接访问 GEO DataSets。  ③ 输入关键词进行搜索  在GEO DataSets页面的搜索框中输入关键词：Solanum lycopersicum growth stages。  点击“Search”按钮进行搜索。  ④ 筛选搜索结果  在搜索结果页面浏览与番茄不同生长阶段相关的基因表达数据集。根据需要，可以使用页面右侧的筛选条件进一步筛选结果（如物种、平台类型、数据类型等）。  点击与实验目标最相关的数据集，进入具体数据集页面查看详细信息。  ⑤ 查看数据集信息  在数据集页面中，查看实验设计、样本描述、测序平台等信息，确保所选数据集包含番茄不同生长阶段的基因表达数据。  点击“Full Dataset Record”查看数据集的完整记录。  ⑥ 下载数据  如果数据集符合需求，点击页面中的“Download”按钮下载基因表达数据。  ⑦ 分析基因表达谱  使用合适的生物信息学工具对下载的数据进行基因表达分析。  可以根据实验需求进行差异表达分析，找到在不同生长阶段显著差异表达的基因。  **Ⅳ 高通量测序数据检索与获取**  ① 打开浏览器并访问NCBI SRA（Sequence Read Archive）主页  URL: https://www.ncbi.nlm.nih.gov/sra  ② 进入SRA数据库的搜索页面  在SRA主页的搜索框中，输入关键词：Glycine max drought stress。  点击“Search”按钮进行搜索。  ③ 筛选搜索结果  在搜索结果页面查看与大豆干旱胁迫相关的转录组数据。可以在页面右侧的“Filters”中使用筛选条件，例如数据类型（如“Transcriptome”）、实验类型（如“RNA-Seq”）、物种等，以帮助缩小搜索范围。  ⑤ 点击相关数据集的标题，进入具体的SRA项目页面。  查看数据集信息  在数据集页面，查看数据集的详细信息，包括实验设计、样本描述、测序平台、文献参考等，以确保该数据集符合实验目标。  ⑥ 下载数据  一般使用命令行工具SRA Toolkit下载大规模数据集。在命令行中使用prefetch或fastq-dump命令下载SRA数据。  **Ⅴ 基因组数据检索**  ① 打开浏览器并访问NCBI主页  URL: https://www.ncbi.nlm.nih.gov/  ② 进入NCBI Genome数据库  在NCBI主页的搜索框中，点击左侧的下拉菜单，选择“Genome”数据库。  ③ 输入关键词进行搜索  在搜索框中输入关键词：Fusarium graminearum。  点击“Search”按钮进行搜索。  ④ 筛选搜索结果  在搜索结果页面中，浏览与Fusarium graminearum相关的基因组信息。一般情况下，搜索结果会显示多个基因组版本，可能包括参考基因组、不同菌株的基因组等。  点击相关的基因组名称，进入详细页面查看基因组的完整信息。  ⑥ 查看基因组信息  在基因组详细页面，查看该基因组的组装版本、注释信息、序列数据等。  通常页面上会显示多个链接，如“FASTA”，“Annotation Release”等。点击“FASTA”链接查看基因组的完整核苷酸序列，点击“Annotation Release”查看基因注释信息。  ⑦ 下载基因组序列  在页面右侧选择“Download Assembly”以下载整个基因组组装文件，包括序列和注释信息。  ⑧ 记录相关信息  记录该基因组的版本号、组装ID、基因组长度、物种注释信息以及下载链接等重要信息。  **Ⅵ UniProt数据库检索**  ① 打开浏览器并访问UniProt主页  URL: https://www.uniprot.org/  ② 进入UniProt搜索页面  在主页的搜索框中输入关键词：Lox soybean 或 Lipoxygenase Glycine max。  ③ 筛选搜索结果  在结果页面浏览与Lox蛋白相关的条目。根据需要，可以使用页面左侧的筛选选项来进一步精确检索，例如：  Organism（物种）：选择 Glycine max。  Reviewed（审查状态）：选择“Reviewed”以查看经过专家审查的条目。  ④ 选择相关蛋白条目  在搜索结果中找到与大豆Lox蛋白相关的条目，点击相关蛋白质的名称进入该蛋白的详细信息页面。  ⑥ 查看蛋白质信息  在蛋白详细页面中，查看以下关键信息：  Protein name（蛋白名）：脂氧合酶（Lipoxygenase，Lox）。  Sequence（序列）：查看或下载蛋白质的氨基酸序列。  Function（功能）：了解该蛋白的生物功能、与其他分子或基因的相互作用、参与的代谢途径等。  Pathology & Biotech（病理和生物技术）：该蛋白在生物技术或病理学方面的应用或作用。  Structure（结构）：查看与该蛋白相关的三维结构（如果有），或者链接至PDB数据库进行结构分析。  Publications（文献）：查看与该蛋白相关的科学文献，了解更多研究背景。  ⑦ 下载序列和信息  如需下载蛋白质序列，可以在“Sequence”部分中点击“Download”按钮，选择FASTA格式下载氨基酸序列。  其他信息（如功能描述、结构数据）可以根据需要复制或下载相应的文档。  ⑧ 记录相关信息  记录蛋白质的UniProt编号、蛋白质名称、功能描述、序列长度及其他感兴趣的生物学信息。  **Ⅶ Phytozome数据库检索**  ① 打开浏览器并访问Phytozome主页  URL: https://phytozome-next.jgi.doe.gov/  ② 登录Phytozome账号  如果没有账户，点击“Sign Up”进行注册并创建一个账户。  如果已经有账户，点击“Login”进行登录。下载数据需要登录账户。  ③ 进入Genome数据库  登录成功后，点击主页导航栏中的“Genome”选项，进入基因组浏览页面。  ④ 搜索大豆基因组  在搜索框中输入“Soybean”或“Glycine max”并点击搜索。  在搜索结果中找到Glycine max，点击进入大豆基因组页面。  ⑤ 查看大豆基因组信息  查看以下信息：基因组版本号、基因注释信息、基因组的物理图谱和结构、基因家族、序列同源性和进化关系等。  ⑥ 下载大豆基因组序列  在基因组页面右上角找到“Download”按钮，点击下载大豆的基因组数据。  选择需要下载的文件类型，例如：  Genome assembly（基因组组装序列），通常以FASTA格式提供。  Annotation data（注释数据），通常以GFF或GTF格式提供。  Gene sequences（基因序列），可以选择下载CDS、mRNA或蛋白质序列，通常以FASTA格式提供。  **四、讨论与分析**  **五、评语与成绩**  指导教师：王遂 | | | |

实 验 报 告

|  |  |  |  |
| --- | --- | --- | --- |
| 实验项目名称：**实验二、序列比对** | | | |
| 实验时间 | 2023-2024年  2学期第14周 | 同组人数 | 1 |
| 实验地点 | 成艺楼309 | 试验台号 | 无 |
| **一、实验目的：**  掌握比对软件的原理及比对过程。  **二、实验材料：**  可访问互联网的计算机。  **三、实验主要步骤与结果：**  **Ⅰ 双序列比对**  ① 序列信息：  >AT1G60350.1  ATGGAGGAAGATGCAGCTTTTGATCTACTCAAAGCCGAACTCTTAAACGCAGAAGACGATGCAATAATCTCACGTTATCTGAAGCGTATGGTCGTCAACGGAGACTCATGGCCTGATCACTTCATCGAAGACGCAGACGTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTTAAACCTCGAACAGAGGCTTGTGGTAAAACCGATGGATGTGAAACTGGTTGCTGGAGGATCATGGGTCGTGATAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAGATTCTCAAGTTCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAACTGGAAGCAAGATCATGTGATTTGCAAGATTCGGTTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTTCTACACTACATCAGAATCACTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCCTTTCATCAGATAAGCAATTGGAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTACGTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATGTTTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTAAAACCGATAGATGCATTAATGGTGGTTACTGGAAAATTTTGCACCGTGATAGGCTGATCAAGTCAAAGTCCGGGATAGTTATTGGTTTCAAGAAGGTGTTTAAGTTTCATGAAACGGAGAAAGAAAGATACTTCTGTGGTGGAGAAGATGTGAAGGTAACTTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAGCAGAATAAATTCTTGTGCGTTATCAAGTTTACTTATGATAACTAA  >AT1G60300.1  ATGGAGGACGACGACGCAGCTTATGATCTAATCAAACACGAACTGTTATACTCAGAAGACGAAGTAATAATCTCACGTTATCTGAAGGGTATGGTCGTTAACGGAGATTCTTGGCCAGATCACTTCATCGAAGACGCAAACGTGTTCACCAAGAATCCAGATAAGGTGTTCAATTCTGAGAGACCTAGATTCGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAAACCGATGGATGTGATTCGGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTGGGAAGATTCTAGGGTTCAAGAAGATACTCAAGTTTTGCCTAAAGAGGAAACCTATAGACTACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAACTGGAAGCAAGATCATGTGATTTGCAAAATTCGGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCAAGCATTTCTACACTACATCAGAATCGGTTCTTGAAAATGAGCTGTTGCCATCTTATGGATATTATTTATCCAATACACAAGAGGAGGATGAATTTTATCTGGACGCGATAATGACTTCGGAAGGAAACGAGTGGCCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGCTTGTGGATCTTCAAGATCGGATGTTTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTGAAACTGATAAATGCGATGGTGGTTACTGGAAGATCCTGCACGGTGATAAGCTGATCAAGTCAAATTTCGGAAAGGTCATTGGTTTCAAGAAGGTATTTGAGTTCTATGAAACGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACGGTAACTTGGACTATACAAGAGTATAGGCTTAGCAAAAACGTGAAGCAGAATAAAGTGTTGTGCGTTATCAAGTTGACTTATGATAGATAG  ② 进入Pairwise Sequence Alignment  URL: https://www.ebi.ac.uk/jdispatcher/psa  ③ 选择Global Alignment中的EMBOSS Needle，Sequence type选择DNA  ④ 输入序列  在“Input Sequences”部分，可以选择输入方式：  直接输入：在文本框中粘贴要比对的两条序列（可以是FASTA格式或单行格式）。  文件上传：选择“Upload”选项，将包含序列的文件上传（支持FASTA格式）。  ⑤ 调整参数  点击“More options”尝试调整参数。  ⑥ 运行比对  确认输入序列和设置无误后，点击“Submit”按钮运行比对。  ⑦ 查看比对结果  比对完成后，页面会显示比对结果，包括：  对齐的序列。  相似性得分、E值等统计信息。  可以查看比对的图形表示（如点图）和详细的比对信息。  ⑧ 下载结果  在结果页面，通常会有下载选项，可以选择合适的格式下载比对结果。  **Ⅱ Blast比对**  ① 序列信息  >AT1G60350.1  MEEDAAFDLLKAELLNAEDDAIISRYLKRMVVNGDSWPDHFIEDADVFNKNPNVEFDAESPSFVIVKPRTEACGKTDGCETGCWRIMGRDKPIKSTETVKIQGFKKILKFCLKRKPRGYKRSWVMEEYRLTNNLNWKQDHVICKIRFMFEAEISFLLAKHFYTTSESLPRNELLPAYGFLSSDKQLEDVSYPVTIMTSEGNDWPSYVTNNVYCLHPLELVDLQDRMFNDYGTCIFANKTCGKTDRCINGGYWKILHRDRLIKSKSGIVIGFKKVFKFHETEKERYFCGGEDVKVTWTLEEYRLSVKQNKFLCVIKFTYDN  ② 打开浏览器并访问NCBI BLAST主页  URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi  ③ 选择蛋白质BLAST（blastp）  在BLAST主页，找到“Basic BLAST”选项，然后选择“Protein BLAST”以进行蛋白质序列比对。  ④ 输入蛋白质序列  在“Enter Query Sequence”部分，可以直接粘贴要比对的蛋白质序列（Fasta格式）。  ⑤ 选择数据库  在“Choose Search Set”部分，选择要搜索的数据库：  Non-redundant protein sequences (nr)：最常用的数据库，包含来自多种物种的大量蛋白质序列。  Swiss-Prot：包含经过手工注释的高质量蛋白质序列。  PDB：如果您希望比对蛋白质结构相关的序列，可以选择PDB数据库。  Other databases：可以根据实验需求选择其他特定的数据库。  ⑥ 调整参数（可选）  在“Program Selection”部分，选择合适的比对算法：  blastp (standard)：常用的标准蛋白质比对程序。  blastp (short)：如果输入的是短序列（如肽段），可以选择这个选项。  在“Algorithm Parameters”部分，可以调整E值阈值、最大比对结果数等参数，通常默认设置就足够。  ⑦ 运行比对  确认所有输入无误后，点击页面底部的“BLAST”按钮开始运行比对。  ⑧ 查看比对结果  比对完成后，会显示比对结果页面，包括：  Graphic Summary：显示匹配结果的图形表示。  Descriptions：列出比对结果中的相似序列及其相关信息。  Alignments：显示查询序列和数据库中匹配序列的详细比对信息。  Taxonomy：提供比对序列所属物种的分类信息。  ⑨ 分析结果  查看每条结果的比对得分、E值、相似性等信息。  **四、讨论与分析**  **五、评语与成绩**  指导教师：王遂 | | | |

实 验 报 告

|  |  |  |  |
| --- | --- | --- | --- |
| 实验项目名称：**实验三、核酸序列分析** | | | |
| 实验时间 | 2023-2024年  2学期第15周 | 同组人数 | 1 |
| 实验地点 | 成艺楼309 | 试验台号 | 无 |
| **一、实验目的：**  掌握核酸序列常规分析方法，包括将序列整理成标准的Fasta格式，序列颠换（reverse sequence），反向互补（reverse complement sequence），序列的组成成分分析，将核酸序列翻译成蛋白质序列等。  **二、实验材料：**  可访问互联网的计算机。  **三、实验主要步骤与结果：**  ① 序列信息：  >AT1G60350.1  ATGGAGGAAGATGCAGCTTTTGATCTACTCAAAGCCGAACTCTTAAACGCAGAAGACGATGCAATAATCTCACGTTATCTGAAGCGTATGGTCGTCAACGGAGACTCATGGCCTGATCACTTCATCGAAGACGCAGACGTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTTAAACCTCGAACAGAGGCTTGTGGTAAAACCGATGGATGTGAAACTGGTTGCTGGAGGATCATGGGTCGTGATAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAGATTCTCAAGTTCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAACTGGAAGCAAGATCATGTGATTTGCAAGATTCGGTTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTTCTACACTACATCAGAATCACTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCCTTTCATCAGATAAGCAATTGGAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTACGTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATGTTTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTAAAACCGATAGATGCATTAATGGTGGTTACTGGAAAATTTTGCACCGTGATAGGCTGATCAAGTCAAAGTCCGGGATAGTTATTGGTTTCAAGAAGGTGTTTAAGTTTCATGAAACGGAGAAAGAAAGATACTTCTGTGGTGGAGAAGATGTGAAGGTAACTTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAGCAGAATAAATTCTTGTGCGTTATCAAGTTTACTTATGATAACTAA  ② 打开“序列操作工具箱v2”  URL: http://www.detaibio.com/sms2/index.html  ③ 选择左面“序列操作工具箱”中的“DNA序列清理”，输入除去大于号开头标题行的所有信息，尝试去掉换行符，数字，修改大小写等。  ④ 选择左面“反向互补序列计算”，输入需要反向互补的序列的Fasta格式文件，尝试进行反向互补。  ⑤ 选择左面“DNA分子量计算”、“DNA碱基统计”和“密码子使用频率计算工具”等，分别计算序列的分子量、碱基组成和密码子使用频率等信息。  ⑥ 打开NCBI的ORFfinder页面  URL: https://www.ncbi.nlm.nih.gov/orffinder/  ⑦ 在对话框输入候选序列，默认参数，点击“Submit”  ⑧ 对结果进行保存整理。  **四、讨论与分析**  **五、评语与成绩**  指导教师：王遂 | | | |

本实验报告由东北农业大学教务处编制

教务处网址：http://jwc.neau.edu.cn