

# 浙江大学校级大学生创新创业训练计划项目

## 结题报告

项目编号 X2022069

项目名称 鸭跖草科的系统发育转录组学研究

项目负责人 裴晨宇

院(系)、年级、专业 生命科学学院 2020 级 生物科学

联系电话 18845371768

电子邮件 3200100225@edu.zju.cn

结题日期 2023 年 5 月 5 日

项目类别：创新训练项目 ☐ 创业训练项目 ☐ 创业实践项目 ☐

校级SRTP项目 ☐ 院级SRTP项目 ☐ 科研实践项目 ☐

### 一、项目信息

项目名称		鸭跖草科的系统发育转录组学研究				
立项经费		1200 元		结题时间	2023 年 5 月 5 日	
项目 负责 人	学号	姓名	年级	所在学院（系）、专业	联系电话	电子信箱
	32001002 25	裴晨宇	2020 级	生命科学学院	18845371768	<a href="mailto:3200100225@zju.edu.cn">3200100225@zju.edu.cn</a>
参 与 成 员	32001003 36	苏正昊	2020 级	生物医学工程与仪器科学学院	17300988722	3200100336@zju.edu.cn
指 导 教 师	姓 名	职 称	所 在 学 院（系）		联 系 电 话	电子信箱
	李攀	副教授	生命科学学院		13757152017	panli@zju.edu.cn

### 二、项目主要内容

成果简介（不少于 500 字）

#### 一、样本采集与数据获取

本研究共计采集了 19 个鸭跖草科样本，几乎覆盖了鸭跖草科的所有亚族，并成功获得其转录组的测序数据。此外，还有 1 个来自 NCBI 的外类群转录组数据和 1 个鸭跖草科转录组数据也合并入数据集中进行分析。

#### 二、质体蛋白编码基因与直系同源单拷贝核基因的获取

本研究共计组装了 20 个转录本数据集，其中 19 个来自采集的样本数据，1 个来自在 NCBI 上公开发表的转录组数据。在上述数据中，我们鉴定出了 1101 个拷贝数为 0/1，且物种覆盖度达 80% 以上的直系同源核基因，通过以上这些核基因来推断鸭跖草科各属之间的系统发育关系。最终的转录组数据集包括 20 个内类群（鸭跖草科）和 1 个外类群（钵子草科）。

同时，我们也从转录本数据集中分离出 19 个质体数据，并将其组装成环，将其蛋白质编码基因进行注释。另外，将 28 个发表在 NCBI 上的鸭跖草科物种的叶绿体基因组合并入数据集中进行分析，对其蛋白质编码基因的 RNA 编辑位点进行预测并手动修改后，与转录组中获得的质体数据合并，共同进行鸭跖草科的系统发育关系的构建。

#### 三、系统发育分析

通过对直系同源单拷贝基因与质体蛋白编码基因的系统发育分析，本研究利用串联法和溯祖法构建了鸭跖草科的系统发育树，解析了鸭跖草科内部的系统发育关系。

### 三、项目实施情况总结（1500 字左右）

#### 项目研究创新点与学术价值

项目研究创新点：此前对于鸭跖草科系统发育学的研究大多是基于质体基因片段或叶绿体全基因组，其具有进化意义的信息位点有限，对于系统发育关系的解析存在局限性，而该研究项目运用较为经济的转录组测序方法，且选取具有代表性的鸭跖草科物种，以较低的成本较为全面地研究了鸭跖草科内部的系统发育关系。

学术价值：本研究利用转录组数据对鸭跖草科内部的系统发育关系进行了比较全面的解析，有助于解释鸭跖草科的真实进化历程。

#### 项目执行过程与实施情况

2022.3-2022.5 文献阅读与基础知识学习：在此阶段我们开始参加导师组织的本科生组会，每学期进行文献阅读的汇报，并且每周学习进化生物学与系统发育学知识，阅读《Tree thinking》《分子生态学》等数目，并开始熟悉基于 Linux 系统的基本操作。

2022.6-2022.8 撰写研究方案并学习分析软件：在此阶段，我们开始着手构思具体研究方法流程，确定数据分析所需要的样本的最低要求有哪些，并且每周学习相应基于 Linux 系统的生信软件，如 Hybpiper, samtools, bowtie2 等，以及阅读大量相关文献学习经典的系统发育学分析流程，如比较基因组学分析流程、系统发育树不同构建方法的操作流程等等。

2022.9-2023.1 数据采集：在这一阶段，我们最终确定了需要采集的物种样本名录，在导师的帮助下，顺利将样本测序，得到了所有样本的转录组测序数据。

2023.2-2023.3 数据分析：这一阶段，我们开始着手对测序结果数据进行相应的组装、分析，但由于服务器算力、学业压力、数据获取时间较晚等多方面因素，数据分析进展比较缓慢，很多分析流程被选择性地简化。

2023.3-2023.5 撰写研究报告：在最后的阶段，不断完善上一阶段未能顺利进行或结果较差的数据分析，总结已有的研究结果，并进行相应的研讨与分析，最终撰写研究报告。

本项目执行过程中，所有成员积极参与，但由于学习进度较为缓慢，数据分析速度受各方面限制较大，实际实施过程比最初实施计划稍缓，且分析结果未能达到原有预期。但总体上仍完成了对鸭跖草科内部系统发育关系的研究。

#### 成果应用与推广及社会效益

暂无

#### 研究存在的不足或欠缺，尚需深入研究的问题等。

①本研究样本量较少，转录组数据不能完整地覆盖鸭跖草科内每一个属，不能将属间的系统发育关系完全解析出来，后续可以继续增加采样，争取对鸭跖草科内部的系统发育关系较为完整稳固地解析出来。

②在转录组提取质体数据的过程中，质体数据质量比较低，成环后覆盖度较差，而且 NCBI 质体数据中仍存在部分 RNA 编辑位点未能预测和修改，导致 NCBI 质体数据与转录组提取的质体数据联合分析的效果欠佳，质体构出的系统发育树支持度不高，其拓扑结构存在解析不完全的问题，后续可以调整预测 RNA 编辑位点的参数，以更严格更精确的限制条件对 RNA 编辑位点进行标记和修改。

③囿于形态学数据获取困难，数据分析进度缓慢等因素，本研究未进行鸭跖草科的性状演化历程研究，后续可以从多方面寻找鸭跖草科各属代表物种的关键性状数据，进行性状演化分析，以探究鸭跖草科真实的进化历程。

④由于进度缓慢，本研究没有进行鸭跖草科的多倍化分析，后续可以使用 Ks 分析等分析手段，检测在鸭跖草科进化的过程中是否有全基因组多倍化的发生，如果有，这些多倍化事件给鸭跖草科提供了怎样的进化动力，这些都是可以深入研究的问题。