

景天属 (Sedum) 三个物种质体基因组种间和种下水平比较分析

G14成员：章思佳、臧政屹、余文静、李文洋

报告人：章思佳

2025.1.11

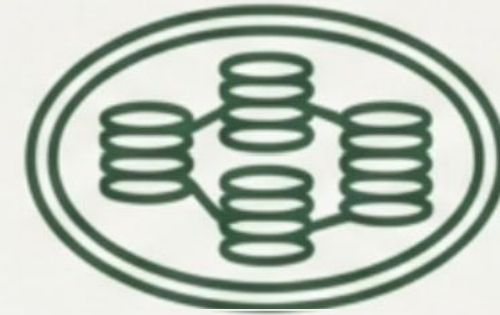
研究背景：为何选择景天属与叶绿体基因组？



核心问题

分类学难题： 景天属 (Sedum) 是景天科中最大的属，包含476个公认物种，但其系统发育关系长期以来模糊不清，并校证实为非单系群 (non-monophyletic)。

数据与分辨率限制： 以往的研究常因数据不足、分辨率有限或分支支持度弱而受限。



选择的方法学路径

叶绿体基因组： 利用叶绿体基因组作为分析工具。

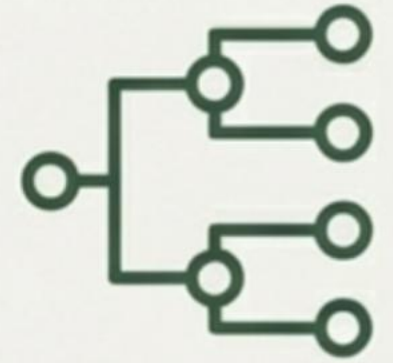
方法优势：

- **母系遗传与低重组率：** 减少了遗传信息的复杂性，提供清晰的遗传谱系。
- **演化速率适中：** 适合用于物种间及物种内的系统发育分析。
- **“超级条形码”：** 其丰富的基因信息使其成为强大的物种鉴定和关系重建工具。

研究目标

1. 揭示三种景天属植物的种内与种间叶绿体基因组变异。
2. 阐明Acre 进化枝的系统发育关系。

从样本到系统树：一个完整的方法学流程



阶段一：数据获取

样本采集与DNA 提取
文库构建与测序

产出

高质量原始测序数据

阶段二：基因组构建

基因组组装
->
基因注释与校正

产出

完整的环状叶绿体
基因组图谱

阶段三：多维度比较分析

结构比较
->
重复序列分析
->密码子分析

产出

揭示不同层级的
基因组变异特征

阶段四：系统发育重建

序列比对与拼接
->
>构建系统发育树

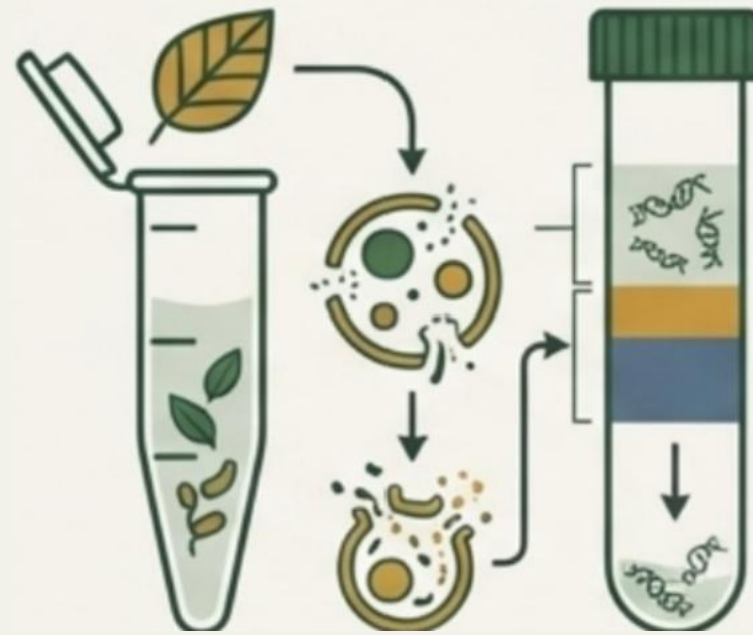
产出

具有高支持度的
演化关系树

方法步骤一：获取高质量的基因组原始数据

分析目标：从16个景天属样本(分属3个物种)中获取纯净、足量的DNA,并将其转化为适合高通量测序的格式。

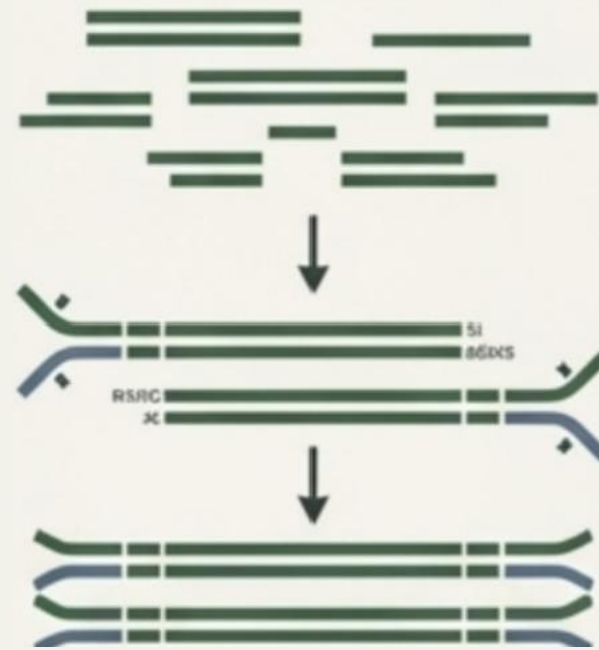
1. DNA提取 (DNA Extraction)



技术：CTAB法 (CTAB method)

原理：使用十六烷基三甲基溴化铵裂解细胞，并通过有机溶剂和酒精沉淀分离纯化DNA。

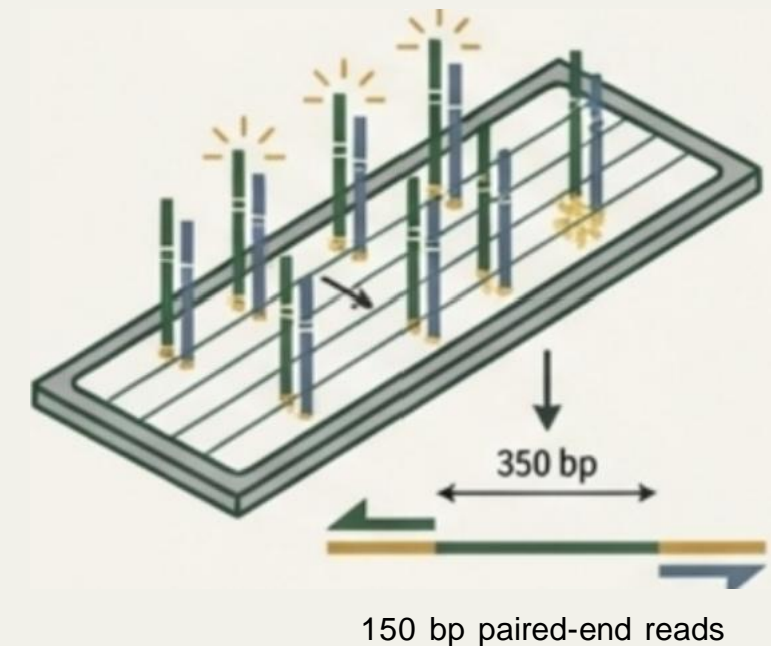
2. 测序文库构建 (Library Preparation)



技术：TruSeq^oDNA PCR-Free Library Prep Kit

方法优势：该试剂盒专为产生高质量、均一的文库而设计，避免了PCR扩增可能引入的偏好性。

3. 高通量测序 (High-Throughput Sequencing)



平台：illumina HiSeqX Ten

参数：150bp双末端测序 (paired-end reads),350bp插入片段大小。

方法优势：Illumina 平台提供高通量和高准确性；双末端测序能提供更精确的序列信息，并有助于后续的基因组组装。

精细解析：IR边界的伸缩与基因变异

分析目标：精确比较不同样本在LSC/IRb, IRb/SSC, SSC/IRa, IRa/LSC四个连接区域的基因分布和长度差异。

工具：IRscope

方法优势：该工具可以直观地可视化基因组结构特征，特别是清晰地展示基因在IR边界上的精确位置和跨界长度，便于进行精细比较。



结果分析：

物种区分标记：ndhF基因延伸至IRb区域的长度可用于区分物种：所有*S. japonicum*样本中该长度为59bp,而其他两个物种为34、44或53bp。

*S. plumbizincicola*在所有基因或区域中表现出最高的长度变异。

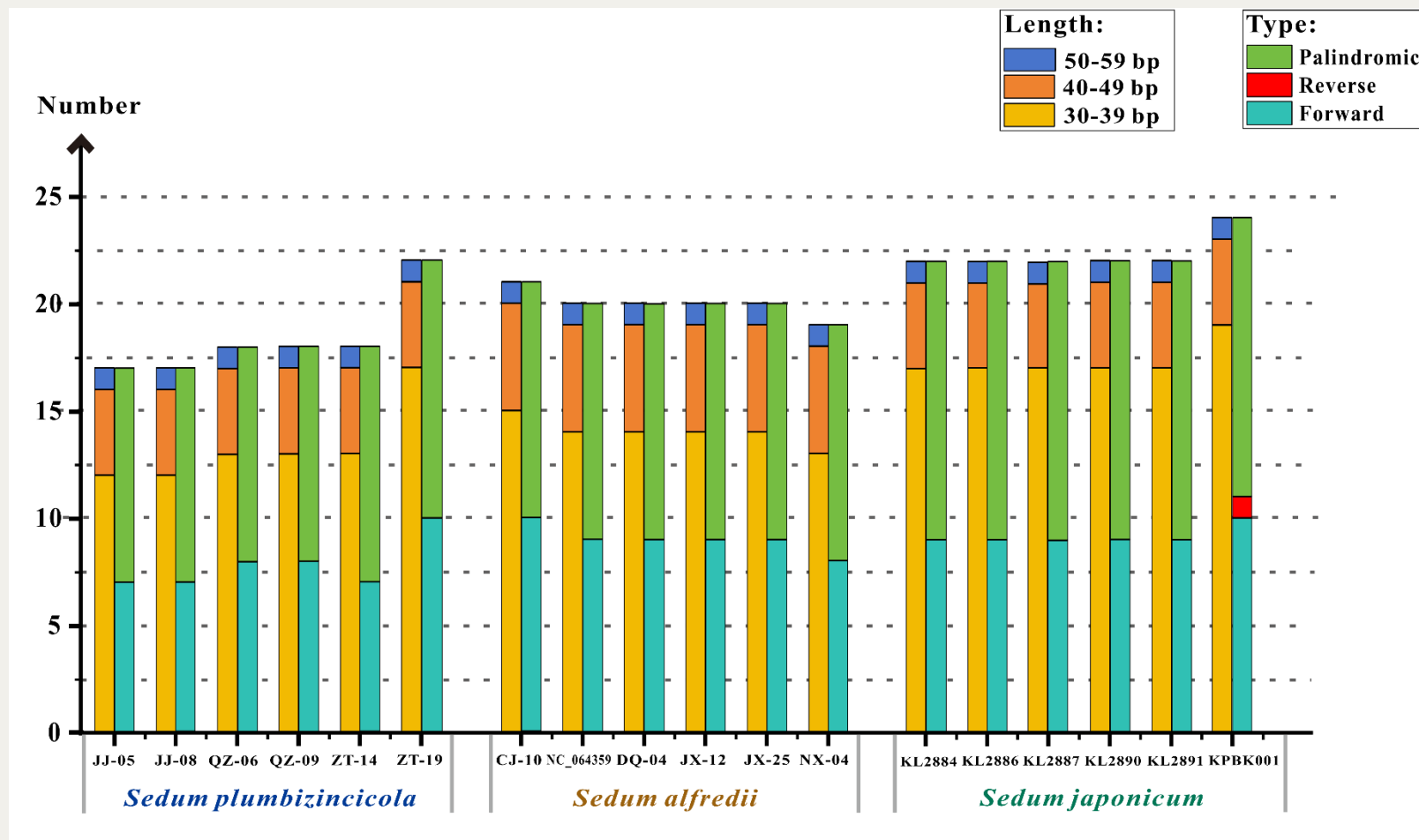
潜在的科级标记：在大多数样本中观察到rps19基因约110 bp的扩张，这可能成为景天科系统发育研究的一个潜在标记。

识别变异热点：重复序列的分布与特征

分析目标

在三个物种的叶绿体基因组中鉴定和比较散在重复序列 (dispersed repeats) 和简单重复序列 (SSRs) 的类型、数量和长度。

散在重复 (Dispersed Repeats):

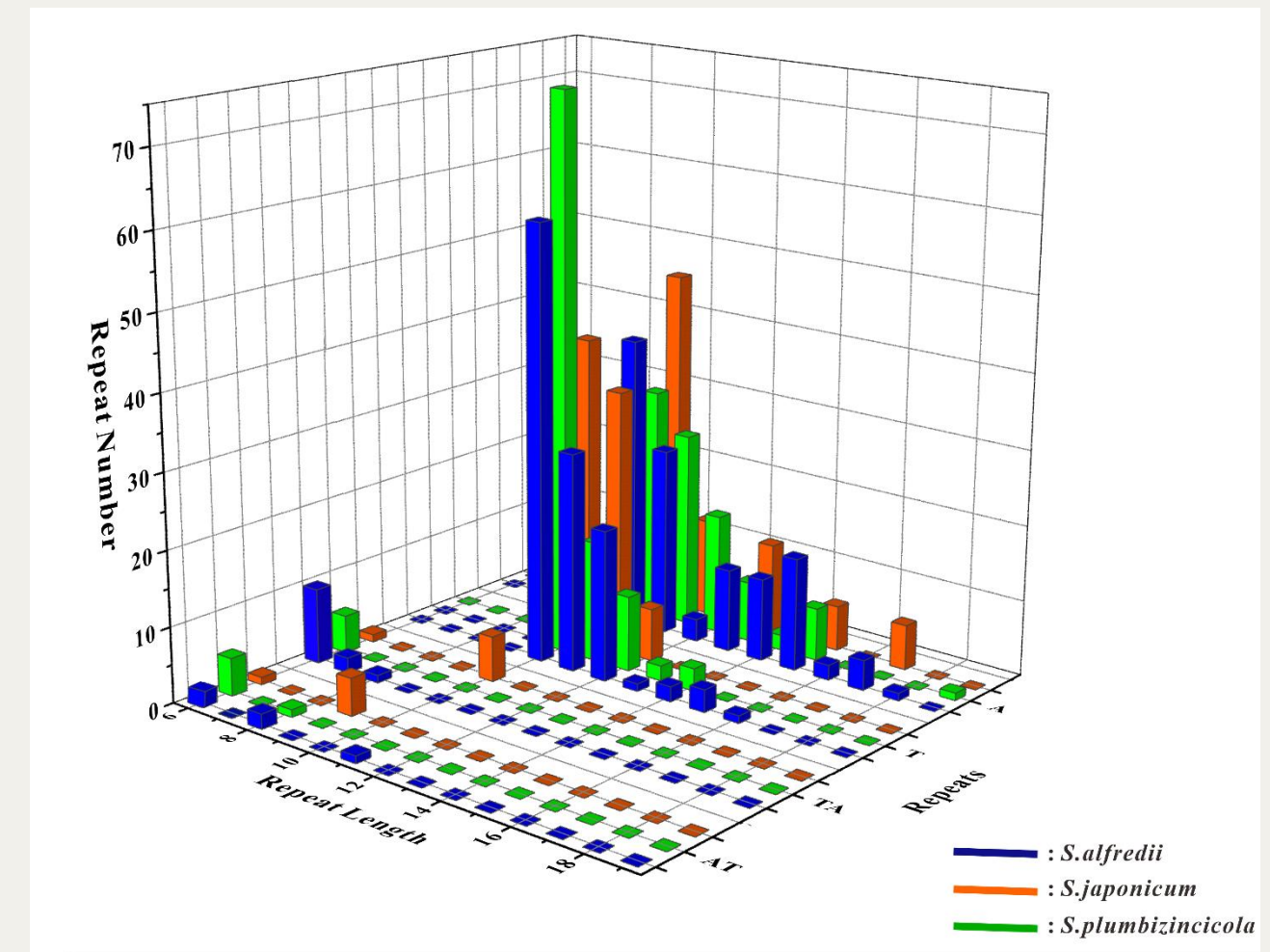


核心方法

1. 散在重复序列分析:
工具: REPuter;
参数设置: hamming distance =3, size limit=30-300 bp。

2. 简单重复序列分析:
工具: Misa;
参数设置: 搜索单核苷酸到六核苷酸基元。

简单重复 (SSRs):



新型标记：利用密码子弃用模式进行物种区分

分析目标：探索在蛋白质编码序列 (CDS) 中完全不被使用的密码子 (即“弃用”密码子)，并评估其作为系统发育标记的潜力。

核心方法：

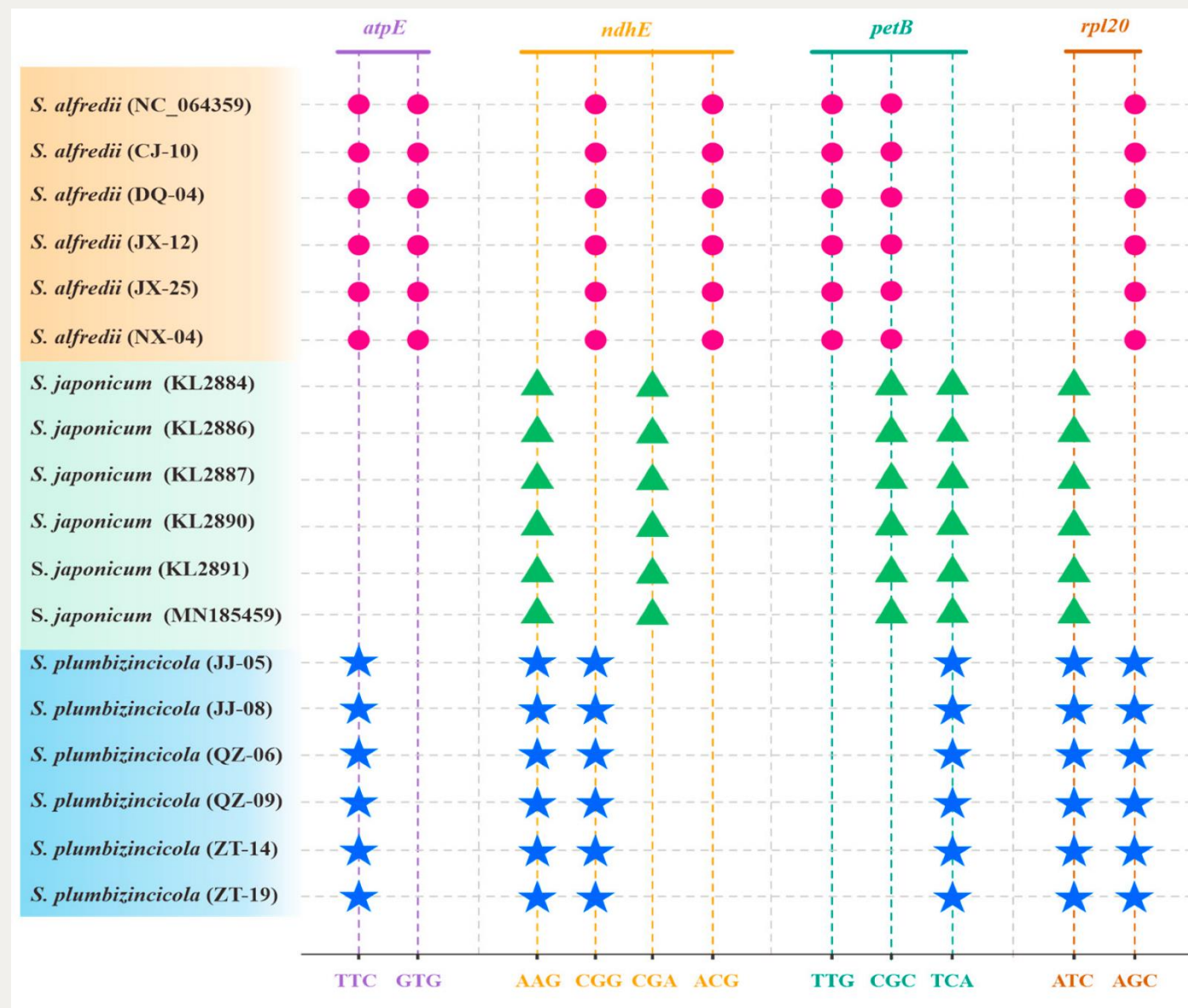
技术：密码子弃用模式 (Codon Aversion Motif, CAM) 分析。

数据：对长度超过300 bp的53个基因进行分析。

结果分析：

强大的物种区分能力：四个基因 (*atpE*, *ndhE*, *petB*, *rp120*) 的 CAMs 可以完全将这三个物种区分开。另有11个基因的CAMs 可以区分 *S.alfredii* 和 *S.plumbizincicola*。

种内变异的存在：在 *S.plumbizincicola* 和 *S.alfredii* 中也观察到个体特异的回避模式，表明该方法在更精细的层面上也具有潜力。



综合证据：构建稳健的系统发育关系

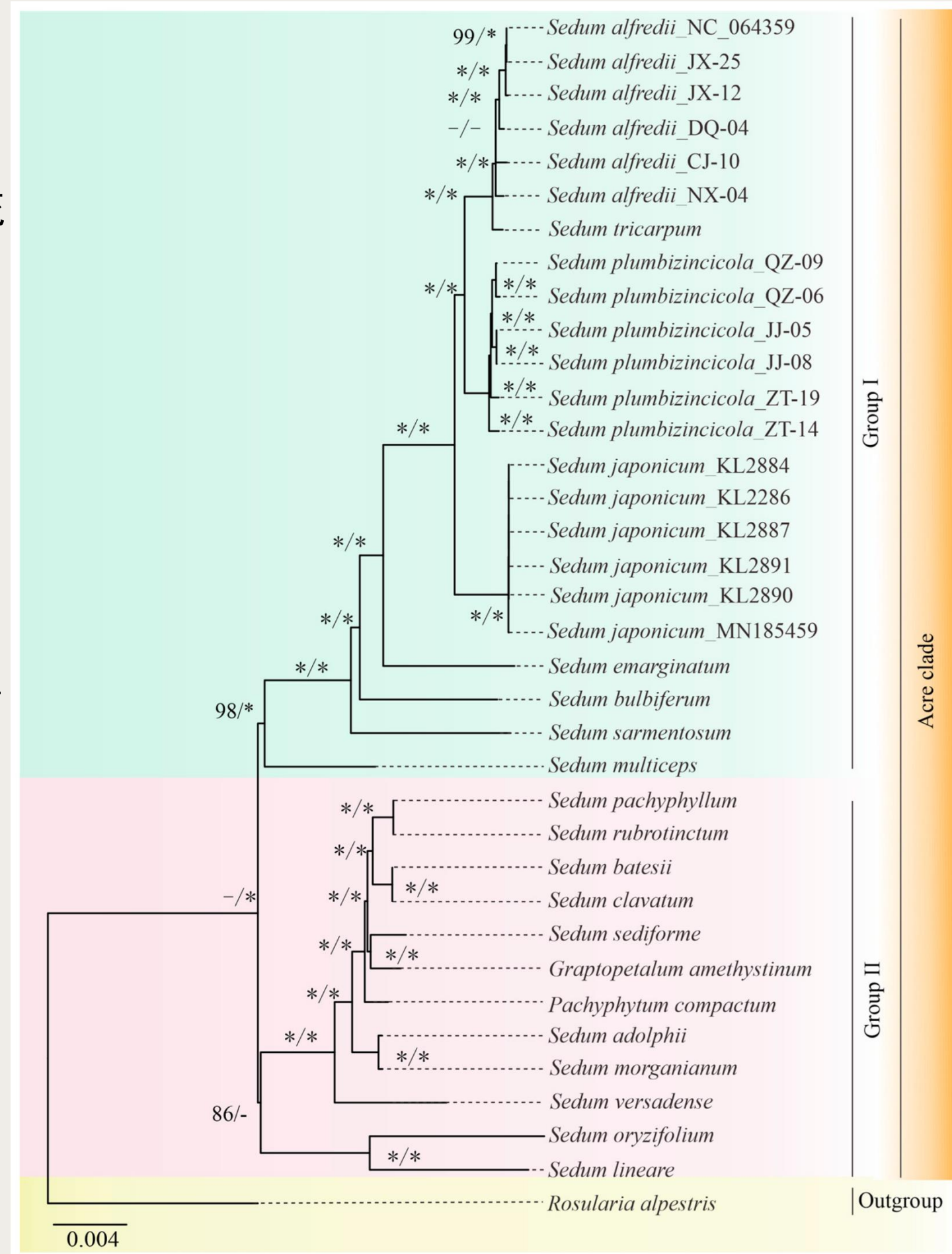
分析目标： 基于79个蛋白质编码基因 (CDS) 的拼接序列，使用两种主流的系统发育推断方法来重建Acre进化枝的演化历史。

核心方法

- **最大似然法 (Maximum Likelihood, ML):** 工具：RAxML8.0.0; 模型：GTRCAT, 进行1000次自举重复 (bootstrap replicates) 以评估分支支持度。
- **贝叶斯推断法 (Bayesian Inference, BI):** 工具：MrBayes 3.2.7; 流程使用ModelTest-NG选择最佳模型，运行1000万代MCMC模拟。
- **方法优势：** 同时使用ML和BI两种不同的算法可以交叉验证结果的可靠性。使用cds数据集(68,382 bp) 能提供丰富的系统发育信号，从而获得高支持度的结果。

结果分析：

- **景天属的非单系性：** 两种方法均支持景天属为非单系群，Acre进化枝可分为两组
- **明确的种间关系：** *S. alfredii*与*S. tricarpum* 关系最近，它们共同与*S. plumbizincicola* 构成姐妹群，且支持度极高 (BS=100, PP=1.00)。
- **清晰的种内聚类：** 同一物种的个体都明确地聚在一起



感谢大家的聆听，敬请批评指正！



G14D: 李文洋

G14C: 余文静

G14A: 章思佳

G14B: 臧政屹

小组成员及分工情况:

章思佳: 数据分析和PPT汇报

臧政屹: PPT整理和美化

余文静: PPT制作和美化

李文洋: PPT制作和美化