Botanical Research 植物学研究, 2022, 11(2), 218-228
Published Online March 2022 in Hans. http://www.hanspub.org/journal/br
https://doi.org/10.12677/br.2022.112026

Hans汉斯

利用叶绿体基因组大单拷贝区的单核苷酸多态位点鉴定紫薇属和马尾藻属植物

索志立[1*]，顾翠花[2*]，左云娟[3]，杨志荣[4]，孙忠民[5]，杨强发[6]，靳晓白[7]

[1]中国科学院植物研究所系统与进化植物学国家重点实验室，北京
[2]浙江农林科技大学风景园林与建筑学院/南方园林植物种质创新与利用国家林业和草原局重点实验室/浙江省园林植物种质创新与利用重点实验室，浙江 杭州
[3]中国科学院东南亚生物多样性研究中心/中国科学院西双版纳热带植物园综合保护中心，云南 勐腊
[4]中国科学院植物研究所植物标本馆，北京
[5]中国科学院海洋研究所海洋生物分类与系统演化实验室，山东 青岛
[6]湖南紫薇投资集团有限公司，湖南 邵阳
[7]中国科学院植物研究所北京植物园，北京

收稿日期：2022年2月18日；录用日期：2022年3月24日；发布日期：2022年3月31日

摘 要

由于植物的种间和属间存在或多或少的形态相似性，用于植物分类的有价值的形态性状数量有限。DNA短片段提供的信息量少，只能区分有限数量的植物种。精确鉴定和数字化管理全球植物，仍然是极其具有挑战性的难题。本文报道，在一个核苷酸分子水平上精确鉴定紫薇属植物的新方法。叶绿体基因组的大单拷贝区的单核苷酸多态位点，作为关键的分子性状用于编制分子分类检索表。为了达到简化和精确鉴定的目的，回避使用如下3类基因组DNA区域的分子性状，即gap区域、poly结构区域及简单重复序列区域。本研究对于紫薇属植物的分类修订、数字化管理、保护和利用具有重要价值。海洋植物大型藻类马尾藻科马尾藻属的9个种/变种也可以用同样的方法成功鉴定，表明我们的方法不仅适用于高等植物的鉴定，也适用于低等植物藻类。

关键词

紫薇属，马尾藻属，叶绿体基因组，大单拷贝区，单核苷酸多态位点，植物鉴定

*通讯作者。

文章引用: 索志立, 顾翠花, 左云娟, 杨志荣, 孙忠民, 杨强发, 靳晓白. 利用叶绿体基因组大单拷贝区的单核苷酸多态位点鉴定紫薇属和马尾藻属植物[J]. 植物学研究, 2022, 11(2): 218-228. DOI: 10.12677/br.2022.112026

A Novel Method for Identification of *Lagerstroemia* and *Sargassum* Taxa Using Single Nucleotide Polymorphic Characters from the Large Single-Copy Region of Complete Chloroplast Genomes

Zhili Suo[1*], Cuihua Gu[2*], Yunjuan Zuo[3], Zhirong Yang[4], Zhongmin Sun[5], Qiangfa Yang[6], Xiaobai Jin[7]

[1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing
[2]College of Landscape and Architecture, Zhejiang Provincial Key Laboratory of Germplasm Innovation and Utilization for Garden Plants, Key Laboratory of National Forestry and Grassland Administration on Germplasm Innovation and Utilization for Southern Garden Plants, Zhejiang A&F University, Hangzhou Zhejiang
[3]Southeast Asia Biodiversity Research Institute of Chinese Academy of Sciences & Center for Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla Yunnan
[4]National Herbarium (PE), Institute of Botany, Chinese Academy of Sciences, Beijing
[5]Department of Marine Organism Taxonomy and Phylogeny, Institute of Oceanology, Chinese Academy of Sciences, Qingdao Shandong
[6]Hunan Ziwei Investment Group Co., Ltd., Shaoyang Hunan
[7]Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing

Received: Feb. 18th, 2022; accepted: Mar. 24th, 2022; published: Mar. 31st, 2022

Abstract

Morphological features to some extent are similar among plant species/genera, so the number of distinguishable morphological traits useful for plant taxonomy is limited. With the progress of global projects of the tree of life, 4 universal DNA barcodes (*rbc*L, *mat*K, *trn*H-*psb*A and ITS) are recommended for global plant classification, 3 (*rbc*L, *mat*K, and *trn*H-*psb*A) of which are from the chloroplast (cp) genome. However, short DNA fragments provide less information and relatively limited number of plant species can be resolved satisfactorily. Assembling and operating the complete cp genome with approximate 150 kb in length require skilled high-level experts and involve a large workload. It is still difficult to accurately identify and digitally manage the global plants. To provide a more convenient, simple and accurate method for plant identification and classification, for the first time, we used single-nucleotide polymorphic characters from the large single-copy (LSC) region of the cp genome for compilation of molecular taxonomic keys to *Lagerstroemia* species. To ensure simplification and accurate identification, we avoided the use of molecular traits from the following three categories of genomic DNA regions: gap regions, poly-N regions and simple sequence repeat regions. With the huge number of plant species in the world, genetic variations, such as those resulted from gene transfer and loss, may cause the target DNA region to be unavailable and further make comparative analysis impossible in plant DNA barcoding in some cases, but the LSC region is present in each cp genome in the plant kingdom. Thus, comparative analysis of all plant species worldwide can be conducted based on sequences from the LSC region. The cp genome can provide rich information for plant identification. Our new methodology is valuable for improving plant taxonomic revision, upgrading the digital management platform and accelerating

phylogenetic and phytogeographic insights into global plants. Nine *Sargassum* species/varieties (Sargassaceae) of marine plants were also successfully identified in the similar way, indicating that our method is suitable not only for identification and classification of higher plants, but also algae of lower plants.

Keywords

***Lagerstroemia*, *Sargassum*, Complete Chloroplast Genome, Large Single-Copy Region, Single-Nucleotide Polymorphic Character, Plant Identification**

1. 引言

过去 200 多年来，全球范围内进行了大规模野外调查、标本采集，根据形态特征进行植物分类，推测其系统发生关系，进行生物地理学研究[1]-[9]。出版了植物志和专著[10] [11]。近年来，中国科学院设立中国－非洲联合研究中心，目的在于协助非洲国家出版和再版植物志[12]。植物志全面记录植物特征，为植物资源的保护和利用提供科学依据。

由于植物的种间和属间存在或多或少的形态相似性，用于植物分类的有价值的形态性状数量有限[3] [13] [14] [15]，这些给植物物种鉴定带来很大挑战。以中国科学院植物研究所植物标本馆为例，该馆馆藏植物标本 295 万份，其中，约有 60 万份标本没有鉴定，原因大多是由于标本缺乏识别特征或者没有相关植物类群的专家来鉴定。这些存疑标本多放置在类群的科尾。Goodwin *et al*. (2015) [3]评价了 21 个国家 40 个标本馆的 4500 份非洲姜科 Zingiberaceae 椒蔻属 *Aframomum* 植物标本，发现至少 58%的标本名称鉴定错误，龙脑香科 Dipterocarpaceae 龙脑香属 *Dipterocarpus* 植物以及旋花科 Convolvulaceae 番薯属 *Ipomoea* 植物的标本也存在类似的情况[3]。热带植物的标本平均有 50%以上名称鉴定错误[3]。植物多样性资源的鉴定和可持续利用已成为世界性重大需求[16]。

近 20 年来，分子性状辅助植物鉴定已经成为热点研究方向之一。叶绿体基因组序列作为一种信息量大、有价值的数据来源用于植物鉴定和系统发生研究已被广泛接受，因其具有相当保守的基因组结构、基因含量和基因顺序[17]-[23]。随着生命之树全球研究计划的进展，推荐使用的 4 个通用 DNA 条形码(*rbc*L, *mat*K, *trn*H-*psb*A 及 ITS)，其中 3 个(*rbc*L, *mat*K, *trn*H-*psb*A)来自叶绿体基因组[2] [24] [25]。由于 DNA 片段提供的信息量少，只能区分有限数量的植物种类，有学者提出，利用叶绿体全基因组序列作为一个超级 DNA 条形码用于植物的物种鉴定[2] [20] [21]。然而，叶绿体基因组序列的全长大约 15 万个核苷酸，操作工作量巨大。为了提供一种更加简便、准确的植物分类鉴定方法，本研究首次利用叶绿体基因组的大单拷贝区的单核苷酸多态位点，研制分子分类检索表，为紫薇属和马尾藻属植物资源的保护与利用提供技术支持。

2. 材料与方法

测序了代表千屈菜科(Lythraceae)紫薇属(*Lagerstroemia*) 16 种[2] [20] [26] [27] [28]的 31 个体的叶绿体全基因组序列(表 1)，利用 MAFFT v7.055b 软件[29] (http://mafft.cbrc.jp/alignment/software)比对。截取其中大单拷贝区的序列，利用 Mega 7.0 软件[30]分析变异位点。从中选择单核苷酸多态位点，利用 Suo *et al*.

(2012, 2015, 2016) [15] [31] [32]描述的方法，研制紫薇属物种的分子分类检索表。检索表中使用的单核苷酸多态位点作为分子数据，利用 Mega 7.0 软件[30]生成 16 种紫薇的系统发生关系树状图。利用马尾藻科(Sargassaceae)马尾藻属(*Sargassum*)的 9 个种和变种的叶绿体基因组序列，验证了本方法鉴定藻类的适用性。

Table 1. Chloroplast genome sequence data of *Lagerstroemia* used in the present study
表 1. 本研究使用的紫薇属植物叶绿体基因组序列数据

Species name 种名	Accession no. 序列号
L. anhuiensis 安徽紫薇	MT019856.1
L. calyculata 副萼紫薇	MT019850.1
L. caudata 尾叶紫薇	MT019857.1
L. excelsa 川黔紫薇	MT019858.1
L. fauriei 屋久岛紫薇	KT358807.1
L. floribunda 多花紫薇	MT019849.1
L. glabra 光紫薇	MT019859.1
L. guilinensis 桂林紫薇	KU885923.1
L. indica 紫薇	MT019862.1
L. intermedia 云南紫薇	MT019847.1
L. limii 福建紫薇	MT019855.1
L. siamica 南洋紫薇	MK881628.1
L. speciosa 大花紫薇	MT019863.1
L. subcostata 南紫薇	KF572029.1
L. tomentosa 绒毛紫薇	MT019851.1
L. villosa 毛紫薇	MT019853.1

1-1a. Type $A_{3200}C_{3205}A_{3473}C_{4484}T_{4510}C_{4728}T_{5126}A_{6345}G_{6681}A_{6698}A_{6882}C_{7186}T_{7188}C_{7416}A_{8170}$ …… Clade I

1-1b. Type $G_{3200}T_{3205}G_{3473}G_{4484}C_{4510}A_{4728}G_{5126}G_{6345}A_{6681}G_{6698}C_{6882}T_{7186}G_{7188}T_{7416}C_{8170}$ …… Clade II, Clade III, and Clade IV

2-1a. Type $C_{6501}C_{7533}A_{8180}G_{18322}T_{23111}C_{49472}C_{66561}$ …… *L. intermedia*

2-1b. Type $A_{6501}A_{7533}G_{8180}T_{18322}C_{23111}A_{49472}T_{66561}$ …… The fifteen species other than *L. intermedia*

2-2a. Type $C_{9614}G_{19138}T_{26998}A_{31946}G_{46643}G_{59457}G_{63996}T_{66298}$ …… *L. siamica*

2-2b. Type $A_{9614}T_{19138}G_{26998}T_{31946}T_{46643}A_{59457}T_{63996}C_{66298}$ …… The fifteen species other than *L. siamica*

2-3a. Type $T_{4640}T_{4805}G_{6183}A_{28722}G_{32302}T_{57428}A_{57598}T_{62004}C_{64003}A_{72939}$ …… *L. speciosa*

2-3b. Type $C_{4640}C_{4805}A_{6183}G_{28722}T_{32302}G_{57428}C_{57598}C_{62004}A_{64003}C_{72939}$ …… The fifteen species other than *L. speciosa*

1-2a. Type $G_{4449}G_{5675}A_{8017}$ …… Clade II
1-2b. Type $A_{4449}T_{5675}C_{8017}$ …… Clade I, Clade III, and Clade IV
2-4a. Type $G_{5449}C_{9040}G_{15021}G_{47734}T_{63700}$ …… *L. fauriei*
2-4b. Type $T_{5449}A_{9040}T_{15021}C_{47734}C_{63700}$ ……The fifteen species other than *L. fauriei*
2-5a. Type $T_{797}C_{18094}A_{27916}A_{43334}T_{59255}G_{63958}G_{70044}$ …… *L. limii*
2-5b. Type $C_{797}T_{18094}C_{27916}C_{43334}C_{59255}C_{63958}T_{70044}$ … The fifteen species other than *L. limii*
2-6a. Type $A_{128}C_{12047}T_{13235}T_{15382}A_{16047}A_{17177}G_{45187}T_{45310}G_{47325}G_{57389}T_{65380}T_{66313}G_{66865}T_{73190}$ …… *L. subcostata*
2-6b. Type $G_{128}T_{12047}G_{13235}G_{15382}C_{16047}C_{17177}T_{45187}C_{45310}C_{47325}T_{57389}G_{65380}C_{66313}A_{66865}C_{73190}$ …… The fifteen species other than *L. subcostata*

1-3a. Type $A_{3084}C_{3321}T_{3629}G_{6275}G_{6888}C_{7192}A_{7193}T_{8212}A_{8301}C_{8308}$ …… Clade III
1-3b. Type $T_{3084}T_{3321}C_{3629}A_{6275}T_{6888}A_{7192}C_{7193}C_{8212}C_{8301}G_{8308}$ … Clade I, Clade II, and Clade IV
2-7a. Type $A_{4489}A_{19018}C_{24365}A_{35296}T_{40206}G_{57701}A_{58874}$ …… *L. calyculata*
2-7b. Type $G_{4489}G_{19018}A_{24365}C_{35296}G_{40206}T_{57701}C_{58874}$ …… The fifteen species other than *L. calyculata*
2-8a. Type $C_{1771}T_{2086}C_{30780}G_{48464}C_{58378}T_{59760}$ …… *L. floribunda*
2-8b. Type $T_{1771}G_{2086}A_{30780}A_{48464}A_{58378}A_{59760}$ … The fifteen species other than *L. floribunda*
2-9a. Type $C_{14273}C_{15084}G_{18548}G_{63558}C_{69056}$ …… *L. tomentosa*
2-9b. Type $T_{14273}A_{15084}A_{18548}A_{63558}A_{69056}$ …… The fifteen species other than *L. tomentosa*
2-10a. Type $A_{6226}C_{6513}C_{7128}A_{12831}T_{17834}A_{65101}$ …… *L. villosa*
2-10b. Type $C_{6226}A_{6513}G_{7128}G_{12831}C_{17834}C_{65101}$ …… The fifteen species other than *L. villosa*

1-4a. Type $G_{3863}C_{4904}A_{5562}G_{5942}G_{6047}G_{8070}$ …… Clade IV
1-4b. Type $T_{3863}A_{4904}C_{5562}T_{5942}T_{6047}C_{8070}$ …… Clade I, Clade II, and Clade III
2-11a. Type $G_{4172}G_{4173}G_{8310}A_{8448}C_{72550}$ …… *L. anhuiensis* and *L. glabra*
2-11b. Type $A_{4172}C_{4173}T_{8310}C_{8448}A_{72550}$ …… The fourteen species other than *L. anhuiensis* and *L. glabra*
3-1a. Type $G_{8981}A_{64031}A_{69912}$ …… *L. anhuiensis*
3-1b. Type $T_{8981}G_{64031}C_{69912}$ …… *L. glabra*
2-12a. Type $G_{6023}A_{6024}G_{13290}G_{26732}T_{35413}T_{47983}C_{59704}C_{67460}$ …… *L. caudata*
2-12b. Type $T_{6023}C_{6024}A_{13290}T_{26732}C_{35413}C_{47983}G_{59704}A_{67460}$ …… The fifteen species other than *L. caudata*
2-13a. Type $T_{7992}T_{28044}G_{37240}T_{41953}T_{55025}A_{57490}T_{58685}C_{72339}$ …… *L. excelsa*
2-13b. Type $C_{7992}A_{28044}T_{37240}C_{41953}G_{55025}T_{57490}G_{58685}A_{72339}$ …… The fifteen species other than *L. excelsa*
2-14a. Type $A_{7894}C_{35671}G_{72262}$ …… *L. guilinensis*
2-14b. Type $T_{7894}T_{35671}T_{72262}$ …… The fifteen species other than *L. guilinensis*
2-15a. Type $T_{3329}A_{4479}T_{4938}C_{8077}G_{30647}G_{66103}G_{67237}$ …… *L. indica*
2-15b. Type $G_{3329}G_{4479}C_{4938}A_{8077}A_{30647}C_{66103}A_{67237}$ …… The fifteen species other than *L. indica*

Figure 1. Molecular taxonomic key of the sixteen species in *Lagerstoemia* based on the single-nucleotide polymorphic characters from the large single-copy (LSC) region in the chloroplast genome. The figure following the nucleotide character indicates the position of the corresponding SNP from the 5' end of the LSC sequence alignment as described previously by Suo *et al.* (2012, 2015, 2016) [15] [31] [32]; the clade numbers correspond to those in Figure 2

图 1. 基于叶绿体基因组大单拷贝区的单核苷酸多态位点的紫薇属 16 种植物分子分类检索表，核苷酸字母后面的数字为单核苷酸多态位点位于叶绿体基因组大单拷贝区内的位置；分支的编号与图 2 中的分支的编号一致

3. 结果

紫薇属 16 个种的叶绿体全基因组序列比对矩阵的长度为 153826 bp。103 个单核苷酸多态位点来自 LSC 区的第 128 个至第 73199 个碱基的区域内，作为分子性状，研制分子分类检索表(图 1)。成功鉴定了紫薇属的物种。利用检索表内使用的 103 个单核苷酸多态位点，生成 16 个物种的系统发生关系树状图(图 2)。16 个物种分为 4 个大的支系，I，II，III 和 IV。使用 3~15 个单核苷酸多态位点可以鉴定出每个分支，利用 3~14 个单核苷酸多态位点可以鉴定出每个物种(图 1)。同一个支系内的物种具有近缘关系(图 2)。分支 IV 揭示，安徽紫薇 *Lagerstroemia anhuiensis* 和光紫薇 *L. glabra* 具有非常近缘的关系，它们共有 5 个独特的单核苷酸多态位点，即 cpDNA_LSC_$\mathbf{G}_{4172}\mathbf{G}_{4173}\mathbf{G}_{8310}\mathbf{A}_{8448}\mathbf{C}_{72550}$。

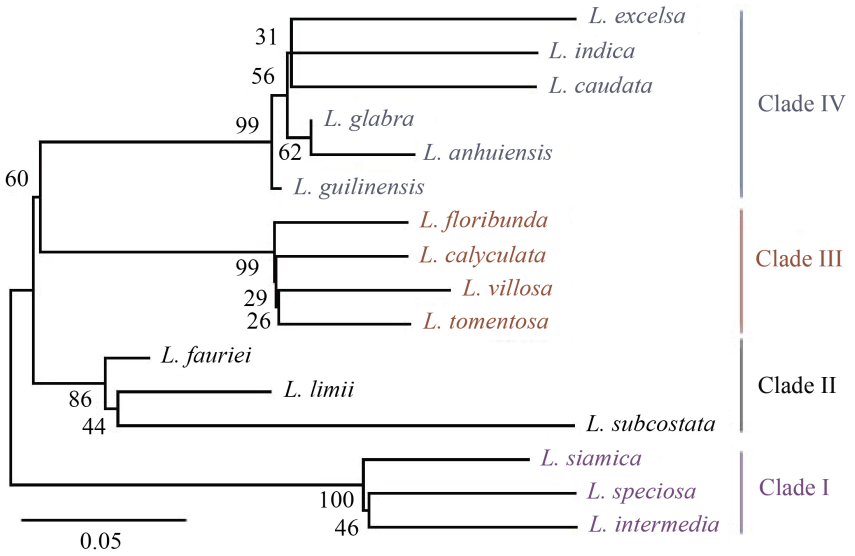

Figure 2. Phylogenetic tree of the 16 *Lagerstroemia* species based on 103 SNPs using the neighbour-joining method with the Tamura-3 parameter model. The numbers near the branches are bootstrap support values (%) of 1000 replications

图 2. 基于 103 个单核苷酸多态位点数据的紫薇属 16 个种的系统发生关系

4. 讨论

紫薇属植物的叶绿体基因组中，大单拷贝区的长度是反向重复区长度的约 3.2 倍，是小单拷贝区长度的近 5 倍。大单拷贝区的变异位点数目，是小单拷贝区的约 2.4 倍，为反向重复区的近 7 倍[20] [26] [27] [28]。大单拷贝区的这 103 个单核苷酸多态位点，已经足够区分 16 种紫薇属植物[27] [28] [33] [34] [35]。紫薇属其余物种的叶绿体基因组测序需要全球合作。

有学者认为 gap 的有无、poly 结构的长度变异，如(A)n，(C)n，(G)n，(T)n，以及简单重复序列的长度变异，可以作为分子性状使用。这 3 类“分子性状”也确实能够用来区分物种。然而，当几个 gap 同时集中存在于临近位置时，即使手动调整，gap 之间的短序列有时也很难匹配到唯一的位置。当 poly 结构长度或简单重复序列的长度很长的时候，容易出现测序误差。在本研究中回避使用 gap、poly 结构以及简单重复序列区域的生物信息。这 3 类“分子特征”不便于编写分子分类检索表。图 2 是 LSC 区的单核苷酸多态位点揭示的系统发生关系。

泛素－蛋白酶体系统的相关基因区域是开发植物分子鉴定工具的另一种选择，精度也较高[17] [31] [32]。然而，获得长度较长的泛素－蛋白酶体系统的相关基因序列，有待实验技术的突破。相对而言，叶

绿体基因组序列的性价比较高，目前更容易获得。

叶绿体 DNA 也用于藻类物种的分类和系统发生关系研究[36] [37] [38] [39]。马尾藻科马尾藻属植物是“海底森林”的重要构成物种。我们利用马尾藻属 9 种/变种的 14 条叶绿体基因组序列(表 2)，获得长度为 74409 bp 的 LSC 区的比对序列，第 5336 个碱基至第 10027 个碱基区段的序列的长度为 4692 bp (占 LSC 全长的 6.3%)，其中，检测到有分类价值的单核苷酸多态位点 221 个，成功鉴定了马尾藻属的 9 个种/变种，见图 3 和图 4。结果表明，我们的植物分类鉴定新方法不仅适用于高等植物，也适用于藻类。

Table 2. Chloroplast genome sequence data of nine *Sargassum* species/varieties used in the present study
表 2. 本研究使用的马尾藻属 9 个种/变种的叶绿体基因组序列数据

Species name 种名	Accession no. 序列号
S. confusum 海蒿子	MG459429.1
S. feldmannii isolate 2017030072 费氏马尾藻	MW784167.1
S. fusiforme 羊栖菜	MN794016.1
S. fusiforme isolate fusiforme 羊栖菜	MN121852.1
S. graminifolium isolate 2016030038 草叶马尾藻	MW784163.1
S. hemiphyllum var. *chinense* 半叶马尾藻中国变种	MT873582.1
S. hemiphyllum var. *chinense* voucher Wu202001 半叶马尾藻中国变种	MT800998.1
S. henslowianum isolate 2017050197 亨氏马尾藻	MW784169.1
S. horneri 铜藻	MN265366.1
S. horneri 铜藻	KP881334.1
S. horneri isolate JD 铜藻	MT795188.1
S. horneri isolate SC 铜藻	MT795189.1
S. ilicifolium var. *conduplicatum* 冬青叶马尾藻重缘变种	MW767830.1
S. kjellmanianum 海黍子	OK323194.1
Scytosiphon lomentaria 萱藻(外群对照，萱藻科 Scytosiphonaceae)	MW415415.1

1-1a. Type $G_{5363}G_{5457}G_{5514}G_{5568}C_{5574}G_{5613}G_{5666}G_{5815}G_{5841}T_{5902}A_{5922}T_{6051}A_{6075}G_{6105}C_{6289}$
$C_{6670}C_{6763}G_{6775}G_{6811}G_{6969}G_{7033}G_{7042}C_{7756}T_{7828}C_{7870}A_{7927}T_{8023}C_{8059}T_{8083}A_{8143}C_{8323}$
$A_{8536}G_{8542}T_{8719}C_{8729}G_{8851}C_{8887}G_{9046}C_{9060}C_{9250}A_{9399}A_{9533}G_{9789}C_{9807}G_{9922}A_{9962}$
…… Clade 1 containing only *S. horneri*

1-1b. Type $T_{5363}A_{5457}A_{5514}T_{5568}T_{5574}A_{5613}A_{5666}A_{5815}A_{5841}A_{5902}G_{5922}C_{6051}G_{6075}A_{6105}T_{6289}$
$T_{6670}A_{6763}A_{6775}A_{6811}A_{6969}A_{7033}A_{7042}T_{7756}A_{7828}T_{7870}C_{7927}C_{8023}A_{8059}A_{8083}G_{8143}T_{8323}$
$T_{8536}A_{8542}G_{8719}T_{8729}A_{8851}T_{8887}A_{9046}T_{9060}A_{9250}T_{9399}G_{9533}T_{9789}T_{9807}T_{9922}T_{9962}$
…… Clade 2, Clade 3, and Clade 4

1-2a. Type $G_{5382}C_{5391}T_{5396}C_{5397}G_{5415}G_{5622}C_{5662}G_{5766}A_{5900}G_{5901}G_{6283}G_{6304}T_{6526}A_{6814}C_{6826}$
$A_{6915}G_{6943}G_{6982}T_{7096}C_{7837}T_{7903}T_{7970}G_{8056}C_{8167}G_{8415}C_{8566}T_{8623}C_{8866}G_{8893}C_{9238}$
$C_{9500}C_{9752}C_{9788}C_{9790}T_{9863}C_{9881}$ …… Clade 2 containing only *S. fusiforme*

1-2b. Type $A_{5382}T_{5391}C_{5396}T_{5397}A_{5415}A_{5622}T_{5662}A_{5766}C_{5900}A_{5901}A_{6283}A_{6304}C_{6526}G_{6814}T_{6826}$
$C_{6915}A_{6943}A_{6982}C_{7096}T_{7837}C_{7903}C_{7970}A_{8056}T_{8167}A_{8415}T_{8566}G_{8623}T_{8866}A_{8893}T_{9238}$
$T_{9500}T_{9752}A_{9788}T_{9790}C_{9863}T_{9881}$ …… Clade 1, Clade 3, and Clade 4

1-3a. Type $T_{5394}G_{5446}A_{6721}A_{7015}T_{8518}G_{8533}$ …… Clade 3
1-3b. Type $A_{5394}A_{5446}G_{6721}G_{7015}C_{8518}T_{8533}$ …… Clade 1, Clade 2, and Clade 4
2-1a. Type $C_{5424}G_{5575}T_{5840}C_{6069}T_{6469}A_{6515}G_{6604}T_{6607}C_{6970}A_{7876}C_{7936}G_{8020}A_{8530}G_{8602}G_{8725}A_{9417}C_{9524}C_{9566}C_{9572}$ …… *S. hemiphyllum* var. *chinense*
2-1b. Type $T_{5424}C_{5575}A_{5840}T_{6069}A_{6469}G_{6515}A_{6604}A_{6607}T_{6970}C_{7876}T_{7936}A_{8020}G_{8530}C_{8602}A_{8725}G_{9417}T_{9524}T_{9566}T_{9572}$
…… The eight taxa other than *S. hemiphyllum* var. *chinense*
2-2a. Type $C_{5376}G_{6691}A_{8656}G_{9009}$ …… *S. kjellmanianum*
2-2b. Type $A_{5376}A_{6691}G_{8656}A_{9009}$ … The eight taxa other than *S. kjellmanianum*

2-3a. Type $A_{5505}C_{5529}T_{5762}A_{5889}T_{5971}G_{6946}C_{8320}C_{8524}T_{9013}$ …… *S. confusum*
2-3b. Type $G_{5505}T_{5529}C_{5762}G_{5889}C_{5971}A_{6946}T_{8320}T_{8524}A_{9013}$
…… The eight taxa other than *S. confusum*

1-4a. Type $C_{5493}G_{5522}A_{5615}C_{5760}T_{6064}T_{6070}C_{6601}G_{6664}A_{6754}A_{6955}G_{8182}T_{8203}T_{8726}T_{9010}C_{9440}A_{9458}G_{9673}C_{9751}T_{10015}$ …… Clade 4
1-4b. Type $T_{5493}A_{5522}C_{5615}A_{5760}C_{6064}C_{6070}T_{6601}A_{6664}G_{6754}G_{6955}A_{8182}C_{8203}C_{8726}C_{9010}T_{9440}G_{9458}T_{9673}T_{9751}C_{10015}$ …… Clade 1, Clade 2, and Clade 3

2-4a. Type $A_{5473}T_{6077}T_{6517}G_{6586}G_{6727}T_{7045}$ …… *S. graminifolium*
2-4b. Type $G_{5473}C_{6077}C_{6517}A_{6586}A_{6727}A_{7045}$
…… The eight taxa other than *S. graminifolium*

2-5a. Type $G_{5336}C_{6108}A_{6246}G_{6532}C_{6598}G_{6910}C_{8176}G_{8464}C_{8704}G_{9182}A_{9754}$
…… *S. ilicifolium* var. *conduplicatum*
2-5b. Type $G_{5336}C_{6108}A_{6246}G_{6532}C_{6598}G_{6910}C_{8176}G_{8464}C_{8704}G_{9182}A_{9754}$
…… The eight taxa other than *S. ilicifolium* var. *conduplicatum*

2-6a. Type $G_{5385}T_{5701}G_{5894}C_{5912}C_{6128}C_{6262}C_{6328}A_{6523}A_{6736}T_{6784}T_{6990}C_{7007}C_{7036}G_{7673}G_{8308}A_{8500}T_{8587}C_{8758}T_{8994}G_{9102}G_{9157}A_{9518}G_{9523}T_{9548}C_{9693}G_{9759}C_{9905}$
…… *S. henslowianum*
2-6b. Type $A_{5385}G_{5701}A_{5894}G_{5912}T_{6128}T_{6262}T_{6328}G_{6523}C_{6736}C_{6784}C_{6990}G_{7007}T_{7036}A_{7673}A_{8308}T_{8500}C_{8587}T_{8758}C_{8994}A_{9102}A_{9157}T_{9518}A_{9523}A_{9548}A_{9693}A_{9759}T_{9905}$
…… The eight taxa other than *S. henslowianum*

2-7a. Type $T_{5423}T_{5640}A_{5691}C_{5704}G_{5896}G_{6076}G_{6136}C_{6251}C_{6689}C_{7054}A_{7082}T_{7551}A_{7940}G_{8512}G_{8761}C_{8819}G_{8835}A_{8845}G_{9286}A_{9437}T_{9556}G_{9925}T_{9969}T_{10027}$
…… *S. feldmannii*
2-7b. Type $G_{5423}C_{5640}G_{5691}T_{5704}A_{5896}A_{6076}A_{6136}A_{6251}T_{6689}T_{7054}C_{7082}G_{7551}C_{7940}A_{8512}T_{8761}A_{8819}C_{8835}G_{8845}A_{9286}G_{9437}G_{9556}A_{9925}G_{9969}A_{10027}$
…… The eight taxa other than *S. feldmannii*

Figure 3. Molecular taxonomic key of the nine species/varieties in *Sargassum* based on the single-nucleotide polymorphic characters from the large single-copy (LSC) region in the chloroplast genome. The figure following the nucleotide character indicates the position of the corresponding SNP from the 5' end of the LSC sequence alignment as described previously by Suo *et al*. (2012, 2015, 2016); the clade numbers correspond to those in Figure 4

图 3. 基于叶绿体基因组大单拷贝区的单核苷酸多态位点的马尾藻属 9 个种/变种的分子分类检索表。核苷酸字母后面的数字为单核苷酸多态位点位于叶绿体基因组大单拷贝区内的位置；分支的编号与图 4 中的分支的编号一致

全球植物的种类数量巨大。基因转移和丢失造成的遗传变异，可能导致靶标 DNA 区域的序列难以获得，进而导致 DNA 条形码研究和系统发生关系研究有时不能进行比较分析[2]，但是，叶绿体基因组(大单拷贝区)在植物界始终存在。因此，基于大单拷贝区的序列，应该能够实现全球所有植物的比较分析。叶绿体基因组可以为植物分类鉴定提供丰富的生物信息。我们的新方法具有确定性好、分辨率高、操作简单、最大限度地减少了多样性管理数据的储存量。对于植物的鉴定、植物数字化管理平台升级以及植物多样性保护，具有重要价值[3] [9]。

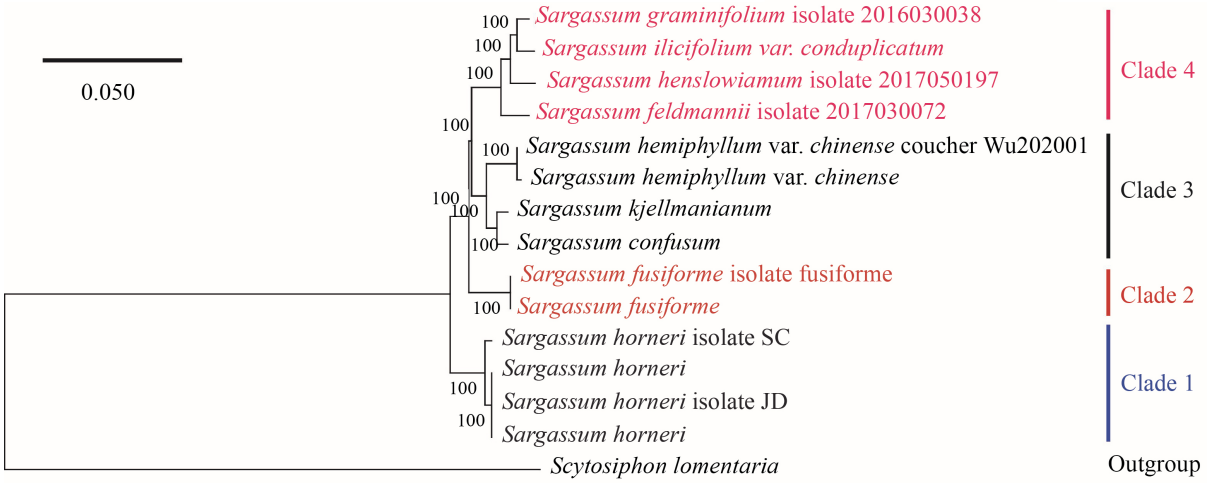

Figure 4. Phylogenetic tree of the nine *Sargassum* species/varieties based on complete chloroplast genome sequences using the neighbour-joining method with the Tamura-3 parameter model. The numbers near the branches are bootstrap support values (%) of 1000 replications

图 4. 基于叶绿体全基因组序列的马尾藻属 9 个种/变种的系统发生关系

致　谢

在样品采集和讨论过程中获得侯伯鑫、张寿洲、余俊杰、徐超、陈进、施济普、宁祖林、徐炳强、张会金、何开红、王少青等多位老师的热情指导和帮助。

基金项目

中国国家自然科学基金项目(No. 31770744)和国家林木种质资源共享服务平台项目(2005DKA21003)。

参考文献

[1] The Angiosperm Phylogeny Group (2016) An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. *Botanical Journal of the Linnean Society*, **181**, 1-20. https://doi.org/10.1111/boj.12385

[2] Zuo, Y.J., Wen, J., and Zhou, S.L. (2017) Intercontinental and Intracontinental Biogeography of the Eastern Asian-Eastern North American Disjunct *Panax* (the Ginseng Genus, Araliaceae), Emphasizing Its Diversification Processes in Eastern. *Molecular Phylogenetics and Evolution*, **117**, 60-74. https://doi.org/10.1016/j.ympev.2017.06.016

[3] Goodwin, Z.A., Harris, D.J., Filer, D., *et al*. (2015) Widespread Mistaken Identity in Tropical Plant Collections. *Current Biology*, **25**, R1066-R1067.

[4] 索志立. 木兰藤科系统位置评述[J]. 植物学通报, 2005, 22(B08): 146-156.

[5] 索志立. 腺齿木科系统位置评述[J]. 植物研究, 2005, 25(1): 26-29. https://doi.org/10.7525/j.issn.1673-5102.2005.01.012

[6] 索志立. 金鱼藻科系统位置评述[J]. 西北植物学报, 2005, 25(5): 1058-1063. https://doi.org/10.3321/j.issn:1000-4025.2005.05.040

[7] 索志立. 水盾草科系统位置评述[J]. 植物学通报, 2006, 23(1): 87-97. https://doi.org/10.3969/j.issn.1674-3466.2006.01.012

[8] 索志立. 莲科系统位置评述[J]. 广西植物, 2007, 27(1): 31-39.

[9] Thomson, S.A., Pyle, R.L., Ahyong, S.T., *et al*. (2018) Taxonomy Based on Science Is Necessary for Global Conservation. *PLoS Biology*, **16**, e2005075.

[10] Wu, Z.Y., Hong, D.Y. and Raven, P.H. (2013) Flora of China, Volume 24. Science Press, Beijing and Missouri Botanical Garden Press. St. Louis.

[11] Muñoz-Rodríguez, P., Carruthers, T., Wood, J.R.I., *et al*. (2019) A Taxonomic Monograph of *Ipomoea* Integrated

across Phylogenetic Scales. *Nature Plants*, **5**, 1136-1144. https://doi.org/10.1038/s41477-019-0535-4

[12] Wei, N., Mutie, F.M., Mwachala, G., *et al*. (2021) *Euphorbia mbuinzauensis*, a New Succulent Species in Kenya from the *Synadenium* Group in *Euphorbia* Sect. *Monadenium* (Euphorbiaceae). *PhytoKeys*, **183**, 21-35. https://doi.org/10.3897/phytokeys.183.70285

[13] 黄建民, 侯伯鑫, 索志立. 邵阳市紫薇品种调查研究 III[J]. 农学学报, 2013, 3(5): 34-41. https://doi.org/10.3969/j.issn.1007-7774.2013.05.010

[14] Suo, Z.L., Zhao, X.Q., Zhao, J.P., *et al*. (2006) 'Sisters on Spring Outing' (*Paeonia suffruticosa* 'Zi Mei You Chun') (Paeoniaceae): A Unique Chinese Tree Peony Cultivar Possessing Side Flowers and Bicolored Floral Discs. *HortScience*, **43**, 532-534. https://doi.org/10.21273/HORTSCI.43.2.532

[15] Suo, Z.L., Zhang, C.H., Zheng, Y.Q., *et al*. (2012) Revealing Genetic Diversity of Tree Peonies at Micro-Evolution Level with Hyper-Variable Chloroplast Markers and Floral Traits. *Plant Cell Reports*, **31**, 2199-2213. https://doi.org/10.1007/s00299-012-1330-0

[16] 洪德元. 生物多样性事业需要科学、可操作的物种概念[J]. 生物多样性, 2016, 24(9): 979-999. https://doi.org/10.17520/biods.2016203

[17] Li, W.Q., Yang, Y., Xie, X.M., *et al*. (2018) *Diospyros oleifera* and *D. deyangensis* Are Revealed as the Closest Relatives to *D. kaki* by E3 Ubiquitin-Protein Ligase UPL3 DNA Sequences. *Hans Journal of Agricultural Sciences*, **8**, 657-673. https://doi.org/10.12677/HJAS.2018.86100

[18] Dong, W.P., Xu, C., Li, D.L., *et al*. (2016) Comparative Analysis of the Complete Chloroplast Genome Sequences in Psammophytic *Haloxylon* Species (Amaranthaceae). *PeerJ*, **4**, e2699. https://doi.org/10.7717/peerj.2699

[19] Dong, W.P., Xu, C., Li, W.Q., *et al*. (2017) Phylogenetic Resolution in *Juglans* Based on Complete Chloroplast Genomes and Nuclear DNA Sequences. *Frontiers in Plant Science*, **8**, Article No. 1148. https://doi.org/10.3389/fpls.2017.01148

[20] Dong, W.P., Xu, C., Liu, Y.L., *et al*. (2021) Chloroplast Phylogenomics and Divergence Times of *Lagerstroemia* (Lythraceae). *BMC Genomics*, **22**, Article No. 434. https://doi.org/10.1186/s12864-021-07769-x

[21] Dong, W.P., Lu, Y.Z., Xu, C., *et al*. (2021) Chloroplast Phylogenomic Insights into the Evolution of *Distylium* (Hamamelidaceae). *BMC Genomics*, **22**, Article No. 293. https://doi.org/10.1186/s12864-021-07590-6

[22] Dong, W.P., Sun, J.H., Liu, Y.L., *et al*. (2021) Phylogenomic Relationships and Species Identification of the Olive Genus *Olea* (Oleaceae). *Journal of Systematics and Evolution*. https://doi.org/10.1111/jse.12802

[23] Dong, W.P., Liu, Y.L., Li, E.Z., *et al*. (2022) Phylogenomics and Biogeography of *Catalpa* (Bignoniaceae) Reveal Incomplete Lineage Sorting and Three Dispersal Events. *Molecular Phylogenetics and Evolution*, **166**, Article ID: 107330. https://doi.org/10.1016/j.ympev.2021.107330

[24] Coissac, E., Hollingsworth, P.M., Lavergne, S., *et al*. (2016) From Barcodes to Genomes: Extending the Concept of DNA Barcoding. *Molecular Ecology*, **25**, 1423-1428. https://doi.org/10.1111/mec.13549

[25] Hollingsworth, P.M., Li, D.Z., van der Bank, M., *et al*. (2016) Telling Plant Species Apart with DNA: From Barcodes to Genomes. *Philosophical Transactions of the Royal Society, London. Series B, Biological Science*, **371**, Article ID: 20150338. https://doi.org/10.1098/rstb.2015.0338

[26] Xu, C., Dong, W.P., Li, W.Q., *et al*. (2017) Comparative Analysis of Six *Lagerstroemia* Complete Chloroplast Genomes. *Frontiers in Plant Science*, **8**, Article No. 15. https://doi.org/10.3389/fpls.2017.00015

[27] Gu, C., Tembrock, L.R., Johnson, N.G., *et al*. (2016) The Complete Plastid Genome of *Lagerstroemia fauriei* and Loss of rpl2 Intron from *Lagerstroemia* (Lythraceae). *PLoS ONE*, **11**, e0150752. https://doi.org/10.1371/journal.pone.0150752

[28] Gu, C., Ma, L., Wu, Z.Q., *et al*. (2019) Comparative Analyses of Chloroplast Genomes from 22 Lythraceae Species: Inferences for Phylogenetic Relationships and Genome Evolution within Myrtales. *BMC Plant Biology*, **19**, Article No. 281. https://doi.org/10.1186/s12870-019-1870-3

[29] Katoh, K. and Standley, D.M. (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, **30**, 772-780. https://doi.org/10.1093/molbev/mst010

[30] Kumar, S., Stecher, G. and Tamura, K. (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution*, **33**, 1870-1874. https://doi.org/10.1093/molbev/msw054

[31] Suo, Z.L., Chen, L.N., Pei, D., *et al*. (2015) A New Nuclear DNA Marker from Ubiquitin Ligase Gene Region for Genetic Diversity Detection of Walnut Germplasm Resources. *Biotechnology Reports*, **5**, 40-45. https://doi.org/10.1016/j.btre.2014.11.003

[32] Suo, Z.L., Li, W.Y., Jin, X.B., *et al*. (2015) A New Nuclear DNA Marker Revealing Both Microsatellite Variations and Single Nucleotide Polymorphic Loci: A Case Study on Classification of Cultivars in *Lagerstroemia indica* L. *Journal of Microbial & Biochemical Technology*, **8**, 266-271.

[33] Furtado, C. and Srisuko, M. (1969) A Revision of *Lagerstroemia* L. (Lythraceae). *Garden Bulletin*, **24**, 185-334

[34] Pham, T.T., Tagane, S., Chhang, P., *et al*. (2017) *Lagerstroemia ruffordii* (Lythraceae), a New Species from Vietnam and Cambodia. *Acta Phytotaxonomica et Geobotanica*, **68**, 175-180.

[35] Qin, H.N. and Graham, S. (2007) *Lagerstroemia*. In: Wu, Z.Y., Raven, P.H. and Hong, D.Y., Eds., *Flora of China*, Science Press, Beijing and Missouri Botanical Garden Press, St. Louis, 277-281.

[36] Huang, C.H., Sun Z.M., Gao D.H., *et al*. (2017) Molecular Analysis of *Sargassum* from the Northern China Seas. *Phytotaxa*, **319**, 71-83. https://doi.org/10.11646/phytotaxa.319.1.3

[37] Zhong, B.J., Xi, Z.X., Goremykin, V.V., *et al*. (2013) Streptophyte Algae and the Origin of Land Plants Revisited Using Heterogeneous Models with Three New Algal Chloroplast Genomes. *Molecular Biology and Evolution*, **31**, 177-183. https://doi.org/10.1093/molbev/mst200

[38] Paudel, Y.P., Hu, Z.X., Khatiwada, J.R., *et al*. (2021) Chloroplast Genome Analysis of *Chrysotila dentata*. *Gene*, **804**, Article ID: 145871. https://doi.org/10.1016/j.gene.2021.145871

[39] Dobrogojski, J., Adamiec, M. and Luciński, R. (2020) The Chloroplast Genome: A Review. *Acta Physiologiae Plantarum*, **42**, Article No. 98. https://doi.org/10.1007/s11738-020-03089-x