

· 专题论坛 ·

## 重要林木樟科植物全基因组测序研究进展

杨智<sup>1, 2, 3</sup>, 杨永<sup>1, 2, 3\*</sup>

<sup>1</sup>南京林业大学南方现代林业协同创新中心, 南京 210037; <sup>2</sup>南京林业大学生物与环境学院, 南京 210037

<sup>3</sup>南京林业大学国家林业和草原局亚热带森林生物多样性保护重点实验室, 南京 210037

**摘要** 近年来, 随着测序技术的革新、测序成本的降低和生物信息学软件的开发, 植物全基因组研究蓬勃发展。樟科(Lauraceae)隶属被子植物木兰类, 泛热带分布, 物种多样性高, 其中很多物种具有重要的经济和生态价值, 目前已发表包括8个物种的13个基因组。该文从樟科全基因组研究现状、基因组特征、起源和进化以及功能基因和基因家族4个方面进行综述, 着重介绍基于组学数据的木兰类及樟科的系统发生、樟科经历的多倍化事件以及与樟科花器官进化和代谢产物相关的基因鉴定。结合研究现状展望了樟科基因组研究的发展方向, 建议通过增加测序基因组分支的代表性并关注具有特殊价值的物种, 及研究物种特异性功能基因以加深对该家族基因功能和进化的理解。

**关键词** 樟科, 演化, 核基因组, 功能基因, 系统发生

杨智, 杨永 (2024). 重要林木樟科植物全基因组测序研究进展. 植物学报 59, 302–318.

近20年来, 基因组研究蓬勃发展(Sun et al., 2022b; Kress et al., 2022)。测序技术和生物信息技术的不断推新升级, 使得基因组、转录组、蛋白质组和代谢组等多组学联合分析成为可能, 极大地推进了植物生物学各领域的研究(谢玲娟等, 2021; Sun et al., 2022a; Kress et al., 2022)。不同植物基因组因其特性不同成为探索特定生物学问题的重要基础(Sun et al., 2022b)。Zhang等(2020)通过对2个榕属(*Ficus*)基因组和1个传粉榕小蜂(*Eupristina verticillata*)的基因组进行比较分析, 探讨了气生根、雌雄同株及雌雄异株的起源, 为揭示动、植物协同进化的分子机制提供了重要资料。藻苔(*Takakia lepidozoides*)、攀枝花苏铁(*Cycas panzhihuaensis*)、油松(*Pinus tabulaeformis*)、百岁兰(*Welwitschia mirabilis*)和无油樟(*Amborella trichopoda*)等基因组的测序不仅为研究陆地植物起源和演化提供了重要依据, 还为进一步研究积累了数据(*Amborella* Genome Project, 2013; Wan et al., 2021; Liu et al., 2022; Niu et al., 2022; Hu et al., 2023b)。通过对水稻(*Oryza sativa*) (Jing et

al., 2023)、葡萄(*Vitis vinifera*) (Dong et al., 2023)和西瓜(*Citrullus lanatus*) (Wu et al., 2023)等农作物的基因组研究发掘出与品质相关的重要基因, 使人们更关注作物的起源和驯化。科学家提出了全部真核生物的测序计划——地球生物基因组计划(The Earth Bio-Genome Project), 目的是提升和重塑人们对生物学、生态系统和进化的理解, 促进生物多样性的保存、保护和再生, 并最大限度地提高基因组资料对社会和人类的回报(Lewin et al., 2018; Blaxter et al., 2022)。

随着测序技术的进步、测序成本的降低以及全球经济的发展, 全基因组研究有加速的趋势。自第1个植物基因组序列被报道(The *Arabidopsis* Genome Initiative, 2000)以来, 截至2020年底, 已有788个植物物种的1 031个基因组被测序和报道(Sun et al., 2022b)。其中农作物的基因组测序受到更多的关注, 已测序物种最多的禾本科(Poaceae)、十字花科(Brassicaceae)和豆科(Fabaceae)均为重要的经济作物科(Sun et al., 2022b)。虽然林木物种基因组研究快

收稿日期: 2023-03-13; 接受日期: 2023-11-14

基金项目: 国家重点研发计划(No.2022YFD2200100)、国家自然科学基金(No.32270217, No.31970205)、江苏省自然科学基金(No.BK20211279)、江苏省研究生科研创新计划(No.KYCX231117)和南京林业大学水杉英才项目

\* 通讯作者。E-mail: yongyang@njfu.edu.cn

速发展,已测序乔木物种达218种,占已测序维管植物的29% (图1A) (谢玲娟等, 2021),而许多木本植物的大科尚未得到足够重视。在全球最大的5个木本植物科中(Beech et al., 2017),仅有桃金娘科(Myrtaceae)已测序24个乔木物种,为已测序乔木最多的科,豆科、茜草科(Rubiaceae)、樟科(Lauraceae)和大戟科(Euphorbiaceae)仅有几个代表物种完成测序(图1B)。

樟科是乔木物种数量排第4的大科(Beech et al., 2017),全世界约有50属, 2 500–3 000种,包含9个主要分支,即樟族(Cinnamomeae)、月桂族(Laureae)、鳄梨族(Perseeae)、桂土楠群(*Mezilaurus* Group)、檬果樟族(Caryodaphnopsidae)、新樟族(Neocinnamomeae)、无根藤族(Cassytheae)、厚壳桂族(Cryptocaryeae)和棠桂族(Hypodaphnideae),其中绝大多数为木本,仅无根藤属(*Cassytha*) 20余种为草质寄生藤本(Rohwer, 1993; The Angiosperm Phylogeny Group, 2016)。樟科是湿润森林中常见的乔木类群,也是亚热带常绿阔叶林的优势类群,具有重要的生态价值(Rohwer, 1993; Liu et al., 2021)。樟科是原始被子植物花演化研究的重要类群,隶属原始被子植物木兰分支樟目(Laurales),具有未分化的原始花被片、花序类型和花性别变异(两性花和单性花) (The Angiosperm Phylogeny Group, 2016; Chen et al.,

2020b)。樟科植物富含多种次生代谢产物,具有特殊气味,使得其成为重要的药品、香料和香水原料(中国科学院中国植物志编辑委员会, 1982; Chen et al., 2020b)。由于其包含特有次生代谢产物,特别是多种芳香族化合物,樟科植物适用于研究植物的化学成分、生态功能和进化关系(Chen et al., 2020a; Xiong et al., 2022)。从基因组水平探讨樟科的进化、花形态演化的分子证据及特殊气味的遗传基础,不仅对揭示樟科植物的生物学特性、开发应用及遗传育种具有指导作用(Chen et al., 2020a),也为阐明被子植物花器官演化、优良木材和芳香族化合物的形成机制提供新见解。

截至2022年底,樟科已有8个代表物种的基因组序列发表,在探讨木兰类系统位置、解决樟科的系统发生问题以及揭示重要代谢产物的分子机制等方面取得了重要进展(Han et al., 2022; Wang et al., 2022)。本文总结了已发表的樟科全基因组测序研究结果,从以下3方面进行阐述。(1) 基于基因组数据的樟科基因组特征研究;(2) 基于樟科组学数据的起源和演化研究;(3) 樟科形态特征、生长特性和代谢产物相关的重要功能基因。此外,还展望了樟科基因组研究的发展方向,以期为林木物种的全基因组研究提供借鉴参考。

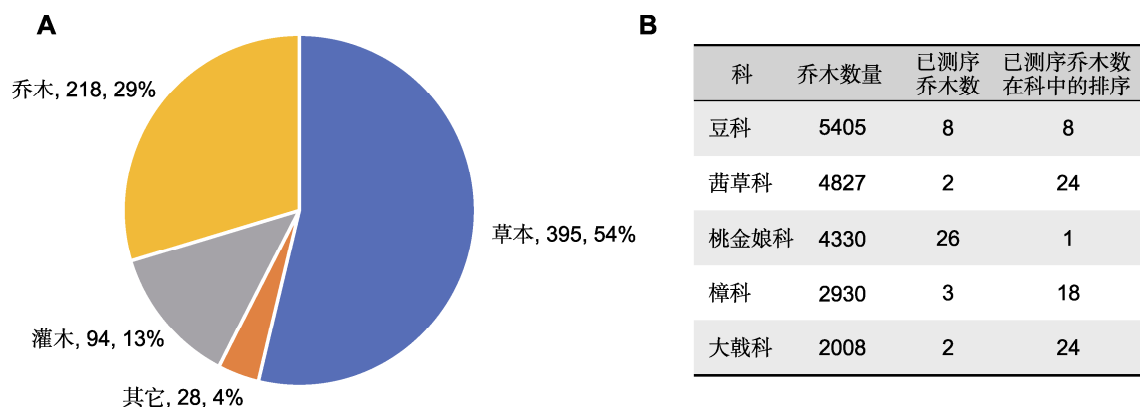


图1 已测序乔木物种概况

(A) 已测序物种的生活型分布,图中依次标注生活型、数量和占已测序维管植物数量的百分比(附录1);(B) 乔木数量最多的5个科测序情况。已测序物种数据来自Sun等(2022b);生活型数据来自GIFT数据库(Weigelt et al., 2020)

Figure 1 Overview of sequenced tree species

(A) The distribution of the growth form of sequenced species, growth form, species number and the percentage of the number of sequenced vascular plants were indicated (Appendix 1); (B) Overview of sequenced top five families with the largest number of tree species. The datasets of sequenced species were obtained from the paper of Sun et al. (2022b); the corresponding datasets of growth form were obtained from the GIFT database (Weigelt et al., 2020)

# 1 樟科全基因组研究进展

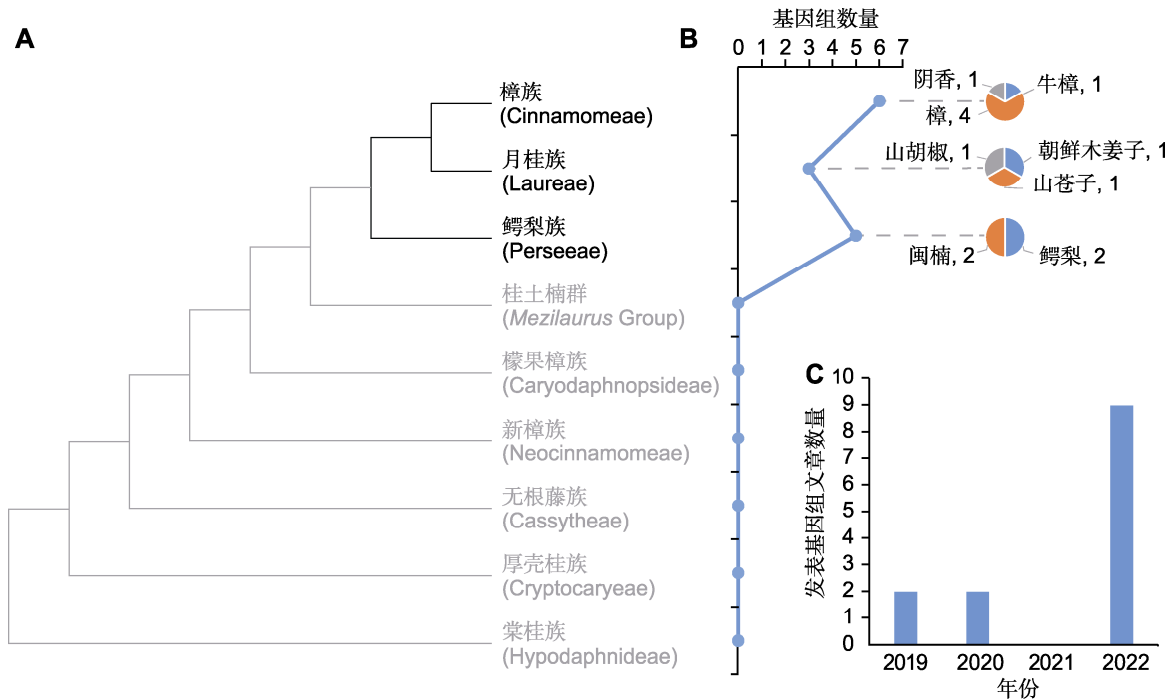
## 1.1 基因组研究概况

自2019年樟科第1个全基因组,即牛樟(*Cinnamomum kanehirae*=*Camphora kanahirae*) (Chung and Hsieh, 2023)的基因组序列发表以来(Chaw et al., 2019),樟科全基因组研究呈现蓬勃发展的态势。目前已有8个物种完成测序(图2;表1),分布于樟科3个族(3/9) 6个属(约6/50)。牛樟基因组不仅是樟科首个完成全基因组测序的物种,也是木兰分支第1个测序的物种(Chaw et al., 2019)。该研究从全基因组层面探讨了木兰类与双子叶植物的关系,发掘了与牛樟形成相关的基因。随后,在2019–2020年,樟科重要的水果鳄梨(*Persea americana*)、化学物质原料山苍子(*Litsea cubeba*)以及优良木材树种闽楠(*Phoebe bournei*)的基因组相继发表(Rendón-Anaya et al., 2019; Chen et al., 2020a, 2020b)。这些物种基因组丰富了樟科基因组数据库,为樟科基因组研究奠定了基础。

2022年是樟科基因组研究成果爆发之年,三代测序技术的发展特别是PacBio公司循环共识测序模式(circular consensus sequencing, CCS)生成HiFi reads (high fidelity reads)的技术促进了樟科全基因组测序研究(Sun et al., 2022a)。仅一年间共有9篇樟科基因组文章发表(图2C;表1),樟(*Camphora officinarum*) (Jiang et al., 2022; Shen et al., 2022; Sun et al., 2022a; Wang et al., 2022)、阴香(*Cinnamomum burmanni*) (Li et al., 2022)、朝鲜木姜子(*Litsea coreana*) (Zhang et al., 2022)及山胡椒(*Lindera glauca*) (Xiong et al., 2022)的全基因组序列相继发布,鳄梨和闽楠的基因组质量得到提升(Han et al., 2022; Nath et al., 2022)。尤其值得关注的是,樟作为我国重要的经济树种,2022年有4个高质量基因组由不同单位主导发表(图2B;表1)。

## 1.2 基因组特征

基因组大小及染色体数目是基因组的基本特征,是研



**图2** 樟科基因组已测序物种概况  
(A) 樟科系统发生树(参考Liu et al., 2021); (B) 樟科物种已测序基因组数量折线图和各族测序物种基因组数量的饼图; (C) 近几年发表樟科基因组文章数量

**Figure 2** Overview of sequenced species of Lauraceae  
(A) Phylogenetic tree of Lauraceae (refer to Liu et al., 2021); (B) Line chart of the number of sequenced genome of Lauraceae, and pie charts of numbers of sequenced species of every tribe; (C) Number of articles on Lauraceae genome in recent years

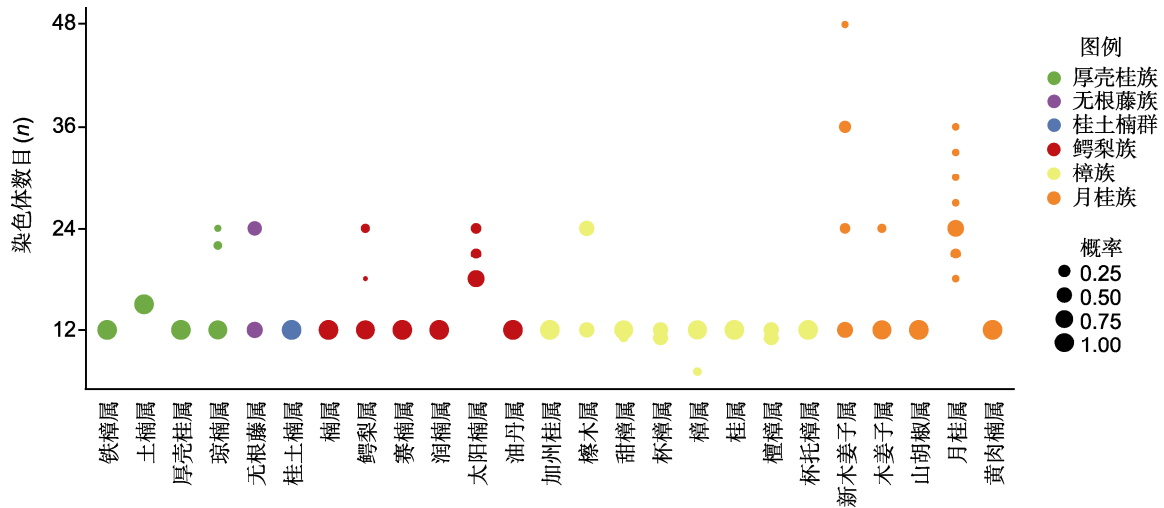
表1 已测序樟科基因组信息

Table 1 Information of sequenced Lauraceae genomes

序号	物种	族	基因组大小 (Mb)	染色体数目 (2n)	杂合度 (%)	编码基因数量	染色体挂载率 (%)	组装完整度 (BUSCO) (%)	注释完整度 (BUSCO) (%)	测序方法	组装水平	参考文献
1	山苍子( <i>Litsea cubeba</i> )	月桂族 (Lauraceae)	1325.7	24	—	31329	94.6	88.4	89.2	PacBio CLR和Hi-C	染色体	Chen et al., 2020b
2	朝鲜木姜子 ( <i>L. coreana</i> )		1139.5	24	1.1	32445	97.1	94.0	—	Illumina、PacBio CCS和Hi-C	染色体	Zhang et al., 2022
3	山胡椒 ( <i>Lindera glauca</i> )		2092.2	24	1.4	65145	94.4	94.2	92.3	Illumina、Nanopore和Hi-C	染色体	Xiong et al., 2022
4	牛樟 ( <i>Cinnamomum kanehirae</i> = <i>Camphora kanahirae</i> )	樟族 (Cinnamomaceae)	730.7	24	—	27899	—	89.0	—	Illumina、PacBio CLR、'Chicago'和Hi-C	染色体	Chaw et al., 2019
5	樟 ( <i>C. camphora</i> = <i>Ca. officinarum</i> )		755.4	24	—	24883	92.4	96.2	—	PacBio CCS和Hi-C	染色体	Jiang et al., 2022
6			723.1	24	1.2	36411	97.9	95.2	90.0	Illumina、PacBio CCS和Hi-C	染色体	Wang et al., 2022
7			719.9	24	2.9	28789	99.9	95.3	89.8	Illumina、PacBio CCS和Hi-C	染色体	Sun et al., 2022b
8			785.0	24	—	29919	85.4	95.2	90.8	PacBio CCS和Hi-C	染色体	Shen et al., 2022
9	阴香( <i>C. burmanni</i> )		1177.6	24	0.7	41549	98.8	89.7	—	Illumina、PacBio CLR和Hi-C	染色体	Li et al., 2022
10	闽楠( <i>Phoebe bournei</i> )	鳄梨族 (Perseeae)	989.2	24	1.5	28198	—	95.0	81.0	PacBio CLR	Scaffold	Chen et al., 2020a
11			941.8	24	1.4	30096	99.2	92.1	91.7	PacBio CLR、BioNano和Hi-C	染色体	Han et al., 2022
12	鳄梨( <i>Persea americana</i> )		912.6	24	—	24616	46.2	85.0	—	PacBio CLR	染色体	Rendón-Anaya et al., 2019
13			913.0	24	—	42769	98.8	98.9	96.6	Illumina和PacBio CCS	染色体	Nath et al., 2022

究科内多倍化和进化的基础。迄今为止, 樟科染色体数目的研究涵盖24个属(约24/50) 136个种(约136/3 000) (Oginuma and Tobe, 2006; Rice et al., 2015)。樟科物种多为二倍体, 通常 $n=12$  (图3; 附录2)。染色体数目变化在樟科属内十分常见, 其中月桂属(*Laurus*)染色体 $n=18, 21, 24, 27, 30, 33, 36$ , 表现出丰富的变异。根据已发表的樟科基因组数据(包括全基因组测序的8个物种) (表1; 附录3), 樟科基因组大小为719 (樟)—2 982 Mb (月桂(*Laurus nobilis*))。月桂族的基因组明显大于樟族和鳄梨族(图4A)。

基因组大小、杂合度和重复序列比例等是判断基因组复杂程度的标准(高胜寒等, 2018)。复杂基因组在测序和组装过程中会面临更大的困难(高胜寒等, 2018)。已测序樟科物种均为二倍体( $2n=24$ )木本植物(表1), 基因组大小介于719–2 092 Mb之间。其中山胡椒基因组是已测序樟科物种中最大的基因组, 约为其它月桂族测序物种的2倍(表1)。樟科物种的基因组杂合度较高(图4B), 除阴香的杂合度为0.7%外, 其余物种的杂合度均大于1%。Sun等(2022a)测序的樟基因组杂合度高达2.9%, 是已测序樟科基因组中杂合



**图3** 樟科各属染色体数目变化  
概率为染色体数目在该属中出现的比例。数据来自染色体数目数据库(CCDB, version 1.66) (Rice et al., 2015)及Oginuma和Tobe (2006) (附录2)

**Figure 3** Variation in chromosome number of various genera in Lauraceae  
Circle sizes represent the percentage occurrence of chromosome numbers across genera. The data was extracted from chromosome counts database (CCDB, version 1.66) (Rice et al., 2015) and result from Oginuma and Tobe (2006) (Appendix 2)

度最高的。樟科基因组重复序列比例为46.1%–76.8%，在樟科基因组中变化较大(图4C)。月桂族重复序列占比最高，超过60%，其中山胡椒的重复序列占比最高，为76.8%。长末端重复序列(long terminal repeat, LTR)为樟科基因组中最常见的转座子(transposable elements, TE)。LTR在樟科已发表基因组中占比为17.0%–52.5% (图4D)，在鳄梨基因组中比例最低，但在闽楠基因组中比例最高(Chen et al., 2020a)。

高杂合度和大量重复序列给樟科基因组测序组装带来了极大困难。最早测序的牛樟、鳄梨、山苍子和闽楠的基因组组装完整度较低(BUSCO组装完整度不超过90%)或者仅组装到Scaffold水平(表1)。二代与三代测序技术结合及生物信息技术的发展极大地提高了全基因组组装质量，得樟科基因组组装完整度和注释基因数量大幅提高。结合Hi-C (highthroughput/resolution chromosome conformation capture) 组装技术和BioNano光学图谱，樟科基因组组装的染色体挂载率得到极大改善，从此前的46.2%–94.6%提升至95%以上(表1)。

1.3 起源和进化

1.3.1 木兰类的系统发生关系

通过基因组测序获取大量的单拷贝核基因数据并构

建生命之树是探讨类群之间系统发生关系的重要方式(王伟和刘阳, 2020)。木兰类、真双子叶植物和单子叶植物之间的进化关系一直困扰着植物学家(Guo et al., 2023)。尽管已有许多木兰类基因组发表，也利用转录组、叶绿体和线粒体数据构建了被子植物系统发生树(One Thousand Plant Transcriptomes Initiative, 2019; Li et al., 2021; Hu et al., 2023a)，但木兰类、真双子叶植物和单子叶植物之间的进化关系仍然存在争议(Shen et al., 2022)。樟科作为木兰类最大的科(The Angiosperm Phylogeny Group, 2016)，木兰类的系统发生位置也是樟科基因组研究的重点内容之一。樟科基因组相关文章中有11篇对该问题进行了讨论(占84.6%)。在樟科基因组研究结果中，木兰类的系统发生位置呈现2种结构(图5B)：I类是真双子叶植物的姐妹群，II类是真双子叶植物和单子叶植物构成分支的姐妹群。其中结构I更常见，得到7篇文章的支持。

在构建系统发生树时，使用不同的采样策略、数据类型、基因数量和分析方法都会导致木兰类系统发生的结构不一致(Chen et al., 2020b; Han et al., 2022; Zhang et al., 2022; Hu et al., 2023a)。已发表的樟科基因组文献中，研究者通常仅使用100–300个单拷贝直系同源基因构建系统发生树，且取样缺少代表性。例如，作为木兰类姐妹群的金粟兰目物种只包



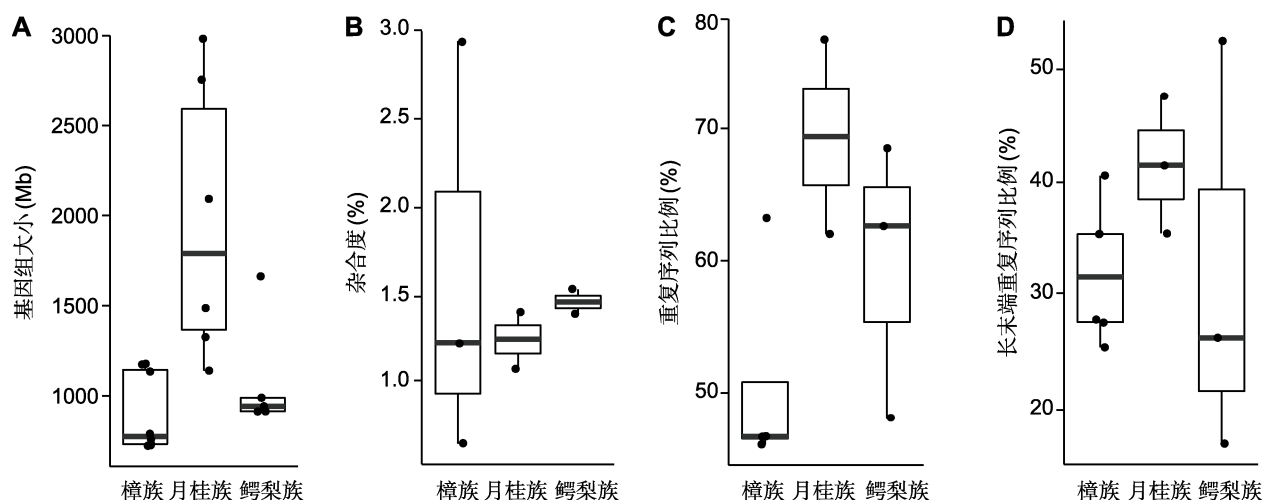


图4 已测序樟科基因组特征比较

(A) 基因组大小; (B) 基因组杂合度; (C) 重复序列比例; (D) 长末端重复序列比例

Figure 4 Comparison of characters of sequenced Lauraceae genomes

(A) Genome sizes; (B) Genomic heterozygosity; (C) Ratio of repeat sequences; (D) Ratio of long terminal repeat

括在Wang等(2022)的基因组研究中。稀疏的分类单元取样和有限的基因数量会影响拓扑结构的可靠性,而不同的数据类型和分析方法也会产生不同的结果。Chen等(2020a)基于核苷酸序列使用贝叶斯法构建的系统发生树以及基于氨基酸序列使用串联法和溯祖法构建的系统发生树均支持拓扑结构I,而基于核苷酸序列使用串联法和溯祖法构建的系统发生树支持拓扑结构II。Zhang等(2022)基于核苷酸和氨基酸序列使用串联法构建的系统发生树支持拓扑结构I,而用溯祖法所得结果支持拓扑结构II。不完全谱系分选(incomplete lineage sorting, ILS)被认为是造成木兰类系统发生关系不一致的重要因素(Rendón-Anaya et al., 2019; Chen et al., 2020a, 2020b; Zhang et al., 2022),但仅有少量研究进行了相关分析。基因家族周转率分析(turnover analysis)表明,在主要被子植物分支同时扩张的过程中存在明显的重复基因更替,因此增加了发生不完全谱系分选的可能性(Rendón-Anaya et al., 2019)。基于溯祖法的ASTRAL分析表明,早期被子植物快速分化时可能发生了不完全谱系分选(Chen et al., 2020a; Zhang et al., 2022)。除系统发生树外,共线性分析可为木兰类的进化关系提供更多证据。在鳄梨基因组研究中, Rendón-Anaya等(2019)基于共线性的系统发生分析,利用位于同源共线性模块上的同源基因序列差异,支持木兰类为单

子叶植物和双子叶植物的姐妹类群。

### 1.3.2 樟科的系统发生关系

自2000年樟科的首篇分子系统学文章发表以来(Rohwer, 2000),科学家为探讨樟科物种的系统发生关系作出了许多努力,但樟科的系统发生研究仍处于初级阶段(Tian et al., 2021)。究其原因主要有2点。(1) 取样代表性不足。Liu等(2021)利用叶绿体基因组构建了樟科目前取样最全且支持率最高的系统发生树,但其取样仅占樟科属的50% (约25/50),种的4% (约129/3 000)。形态性状变异小、形态的平行演化或逆转导致物种分类研究比较困难,以及乔木生长在难以到达的热带森林中,造成调查和采样困难,是影响樟科取样代表性的2个主要因素(Li et al., 2004; 李捷和李锡文, 2004; Tian et al., 2021)。(2) 分子标记信息不足。常用的分子片段,如 $matK$ 、 $trnL$ - $trnF$ 、 $trnT$ - $trnL$ 和 $psbA$ - $trnH$ 的分辨率不足以构建稳定的系统发生树(Rohwer, 2000; Chanderbali et al., 2001; Li et al., 2004, 2008; Rohwer and Rudolph, 2005),而樟科ITS序列存在假基因和不完全一致性进化,造成直接测序获得的序列有严重的杂合现象,无法单独提供可靠的节点推断(黄建峰等, 2016)。虽然通过使用叶绿体基因组确定了樟科的基本框架,但樟科叶绿体基因组十分保守,无法解决各分支内属的系统发生关系

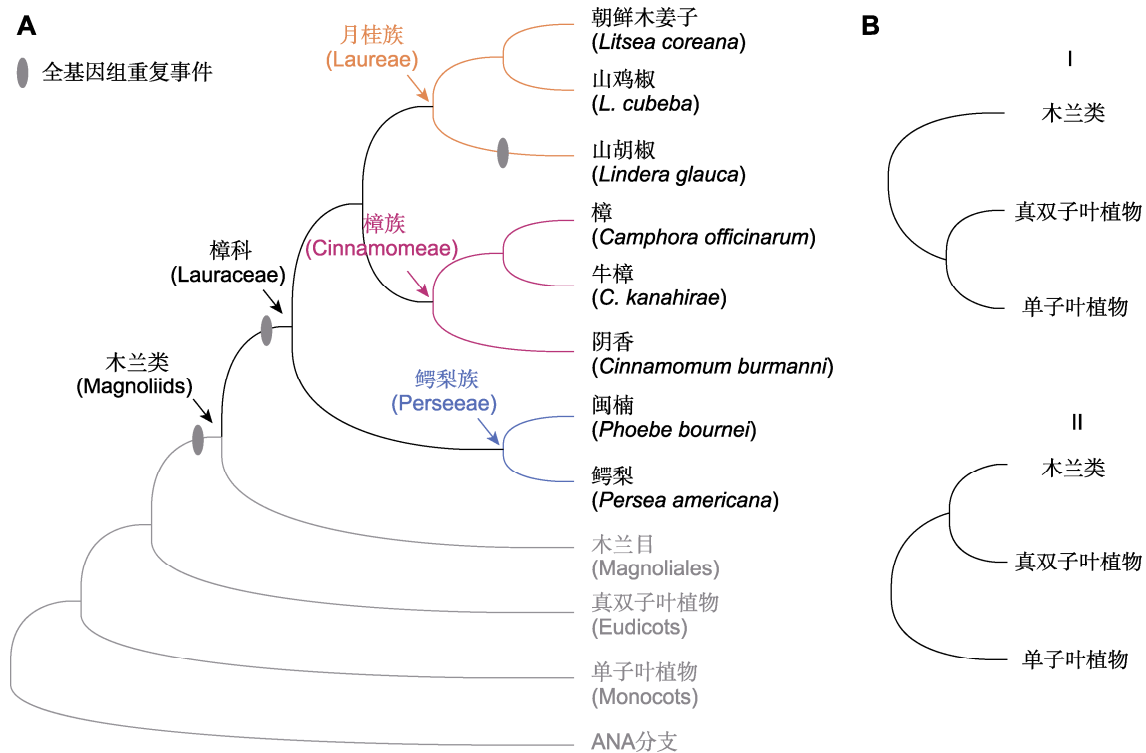


图5 樟科已测序物种系统发生关系

(A) 樟科已测序物种系统发生树(参考Liu et al., 2021; Han et al., 2022) (灰色椭圆形代表樟科经历的全基因组重复事件); (B) 从樟科基因组文献中得到的木兰类、真双子叶植物和单子叶植物系统发生关系的2种拓扑结构。

Figure 5 Phylogenetic relationships of sequenced species of Lauraceae

(A) Phylogenetic tree of sequenced species in Lauraceae (refer to Liu et al., 2021; Han et al., 2022) (The grey ovals represent predicted whole genome duplication events in Lauraceae); (B) Two topologies of the relationship of magnoliids, eudicots and monocots in the published papers based on the studies on Lauraceae genomes

(Xiao et al., 2020, 2022; Trofimov et al., 2022; Xiao and Ge, 2022)。

全基因组研究为樟科多层次系统发生提供了证据和大量的数据, 揭示了樟科复杂的进化历史。在山苍子基因组研究中, 研究人员结合樟科代表属的浅层基因组和转录组数据, 从基因组层面为樟科复杂的系统发育关系提供了新的证据(Chen et al., 2020b)。该研究从樟科转录组数据中鉴定出275个单拷贝核基因, 采用串联法和溯祖法推断樟科的系统发生关系, 构建了樟科19属46种的质体系统发生树。研究显示樟科核基因树与前人构建的质体树有显著差异(Liu et al., 2021), 且不一致的节点存在基因树和物种树的冲突。在质体树中, 厚壳桂属(*Cryptocarya*)是樟科其它类群的姐妹群, 而核基因树中樟科唯一的半寄生藤本类群无根藤属是其它类群的姐妹群(Chen et al., 2020b)。此结果与以往基于核基因和质体分子片段得

到的不一致拓扑结构相似(Rohwer, 2000; Chanderbali et al., 2001; Rohwer and Rudolph, 2005)。在系统发生基因组学中, 不一致的进化历史推断十分普遍。不完全谱系分选、杂交(hybridization)和基因渐渗(gene introgression)等生物因素以及数据处理过程中的随机和系统误差等非生物因素都可能造成系统发生树不一致(王伟和刘阳, 2020; Steenwyk et al., 2023)。樟科系统发生树出现明显的拓扑结构冲突, 强调了杂交起源和网状进化对其类群系统发生关系的重要影响(Liu et al., 2021)。而月桂族、樟族和鳄梨族经历了快速辐射分化, 极有可能受到了不完全谱系分选的影响(Xiao et al., 2020, 2022; Liu et al., 2021; Xiao and Ge, 2022)。

在属级层面, 基因组数据也为确立广义樟属及其属下等级的系统发生关系提供了新见解。在樟的基因组研究方面, 首次使用了重测序方法探究广义樟属

(*Cinnamomum* s.l.)内物种的亲缘关系(Wang et al., 2022)。利用重测序数据获得的单核苷酸多态性位点(single nucleotide polymorphisms sites, SNPs)和插入缺失位点(insertions and deletions sites, InDels)构建的广义樟属24个物种系统发生树显示, 樟属(*Camp-hora*)和桂属(*Cinnamomum*)植物形成两大分支。此外, 通过比较基因组学, Li等(2022)发现相较于牛樟和樟, 染色体结构和LTR的变化使阴香具有更为独立的进化历史。这些结果支持Yang等(2022a)将广义樟属划分为樟属和桂属的处理。

### 1.3.3 樟科的多倍化事件

多倍化(polyploidy)或全基因组重复(whole genome duplication, WGD)事件使基因组内所有基因发生重复, 为生物进化提供了原始遗传材料, 被认为是进化的加速器(Wu et al., 2020; 葛颂, 2022)。多倍化在有花植物进化中尤其频繁, 对其适应性进化产生重要影响(Wu et al., 2020)。

樟科基因组研究中鉴定出3次多倍化事件(图2A)。在樟科所有测序基因组中都检测到木兰目和樟目分化之前发生的1次WGD事件(约118–147 Ma)(Sun et al., 2022a; Han et al., 2022)以及已测序樟科物种共同祖先中发生的1次WGD事件(约76–95 Ma)(Shen et al., 2022; Han et al., 2022)。较近的樟科WGD事件与晚白垩纪早期樟科植物随着冈瓦纳大陆的分裂快速辐射分化的时间一致(Chanderbali et al., 2001)。WGD事件通过改善基本生理活动和初级代谢使植物对环境变化的适应性增强, 促进了早期樟科物种的快速辐射(Chen et al., 2020a; Jiang et al., 2022; Sun et al., 2022a)。值得注意的是, 山胡椒基因组在与相近的山苍子分化后(14.90–23.18 Ma)独自经历了1次多倍化事件, 与早中新世(23 Ma)月桂族快速辐射分化的时间一致(Qin et al., 2023)。月桂族为樟科中雌雄异株类群, 主要分布在亚洲热带和亚热带地区, 为亚热带常绿阔叶林的建群种(Li et al., 2004; Qin et al., 2023)。这次多倍化事件导致山胡椒具有比山苍子及其它樟科植物大将近1倍的基因组和更多的蛋白质编码基因(表1)(Xiong et al., 2022)。其中, 与萜烯合酶生物合成相关的基因家族发生显著扩张, 可能对山胡椒的生态适应性和生物适应性有重要作用(Pichersky and Raguso, 2018; Xiong et al., 2022),

有利于其应对早中新世东亚季风盛行引起的极端天气(Qin et al., 2023)。

## 1.4 功能基因和基因家族

通过基因组测序挖掘植物的重要功能基因、解析其代谢通路并进行功能验证是基因组研究的常见模式(Eckardt et al., 2021)。这些研究为植物进化、发育和功能生物学提供了新见解(Eckardt et al., 2021)。

在樟科基因组中, 花器官进化、木材生长、代谢产物合成和炭疽病抗性等相关基因已得到研究(图6)。通过山苍子基因组结合转录组数据, 发现1个高度保守的参与花序形态发生的基因*FUWA* (*PETAL LOSS*)所构建的花序系统发育与樟科物种的系统进化相对应, 揭示了樟科花序从穗状花序、穗状圆锥花序向聚伞形圆锥花序、假伞形花序, 进而到伞形花序演变的规律(Chen et al., 2020b)。Sun等(2022a)在樟的基因组研究中通过花发育相关基因的表达分析和qRT-PCR验证早期被子植物花器官广泛活跃的祖先ABCDE模型。Nath等(2022)解析了鳄梨基因组中独特的庚糖生物合成途径, 发现其涉及景天庚酮糖1,7-二磷酸(sedoheptuloase 1,7-bisphosphate)这一中间体。在其基因组中还观察到大量内切葡聚糖酶基因, 支持鳄梨具有与纤维素酶有关的果实成熟机制, 这与大多数水果以内聚半乳糖醛酸酶作为糖苷酶和水解酶参与植物组织的浸渍以及软腐烂的机制不同(Wakabayashi, 2000)。Wang等(2022)通过比较樟和其它樟科植物鳄梨、牛樟、闽楠和山苍子的基因组, 筛选出中碳链油脂合成的关键候选基因*FatB1* (终止脂肪酸碳链延伸)、*type-B LPAAT*和*DGAT2b* (参与甘油三酯装配), 并绘制出中链甘油三酯(medium-chain triglyceride)合成途径。

樟科植物基因组中功能基因的鉴定涉及内容丰富, 但整体研究不够深入, 仍停留在基因鉴定和解析代谢通路层面, 仅有萜类化合物相关基因进行了功能验证(图6)。萜类化合物主要包括单萜、倍半萜和不规则萜类, 是结构多样化代谢产物中最大的一类(Christianson, 2017)。它们在植物生理功能中起至关重要的作用, 包括生长发育、环境适应、吸引传粉者以及抗病虫害和高温, 也是花和果实香味形成的重要活性物质(Nagegowda, 2020; Jia et al., 2022)。萜类合酶(terpene synthase, TPS)是植物多样化萜类骨



架形成的关键,将甲基赤藓糖醇-4-磷酸途径(methylerythritol-4-phosphate pathway, MEP)或甲羟戊酸途径(mevalonate pathway, MVA)产物进一步催化生成各种萜烯类化合物(Jia et al., 2022)。樟科植物产生的挥发性有机化合物在香料和香水工业中具有很高的应用价值(中国科学院中国植物志编辑委员会, 1982)。不同樟科树种产生的芳香族化合物成分不同,主要是萜类及其衍生物(Xiong et al., 2022),山苍子、月桂和樟的气味主要由不同的单萜类化合物组成,而楠属物种气味主要为倍半萜类化合物(Joshi et al., 2009; Han et al., 2022)。因此,揭示萜类合酶基因家族重要基因的代谢途径,探究萜类合酶基因家族在樟科中的演化是樟科基因组研究的热点。

基于基因组、转录组和代谢组等多组学数据的联合分析,研究者鉴定出樟科中参与重要萜类化合物合成的关键基因并绘制了合成代谢通路图。通过转录组分析、同源和异源瞬时过表达以及体外酶活实验,鉴定了调控山苍子主要单萜化合物合成的关键基因(Chen et al., 2020b)。其中*LcuDXS3*在果实中高表达,可正调控单萜类和倍半萜类多种成分的含量;*LcuTPS-*

22在叶片中特异性积累,调控樟科植物叶片中主要挥发性成分 $\alpha$ -蒎烯、 $\beta$ -蒎烯、桉油醇、苧烯、桉油醇和樟脑的积累;*LcuTPS42*调控芳樟烯、香叶醇和樟科特定香味的主要成分芳樟醇的生物合成。山胡椒基因组研究揭示了其果实和叶片香气中的主要单一萜烯类化合物、 $\beta$ -罗勒烯与D-吉玛烯合成代谢过程中关键酶基因的变化规律,绘制出合成代谢通路图(Xiong et al., 2022)。Wang等(2022)通过比较樟的5种化学型叶片转录组,筛选差异表达的*TPS*基因,并通过异源表达-体外催化反应与亚细胞定位等实验验证了*CcTPS16*和*CcTPS54*分别调控1,8-桉叶油素和异橙花叔醇的合成。Li等(2022)通过阴香的比较基因组学和2种化学型的转录组学及基因克隆和qPCR分析,确定了右旋龙脑合成的2条关键分子途径及相关的功能基因,推测MEP途径可能是阴香合成右旋龙脑的关键途径,而*CbTPS-1*可能是导致阴香叶片右旋龙脑高产的关键调控基因之一。

萜类化合物不仅参与樟科植物的化学组成,还参与抗病机制。在闽楠基因组8号染色体上,Han等(2022)鉴定到特异的基因串联重复事件,其中涉及9

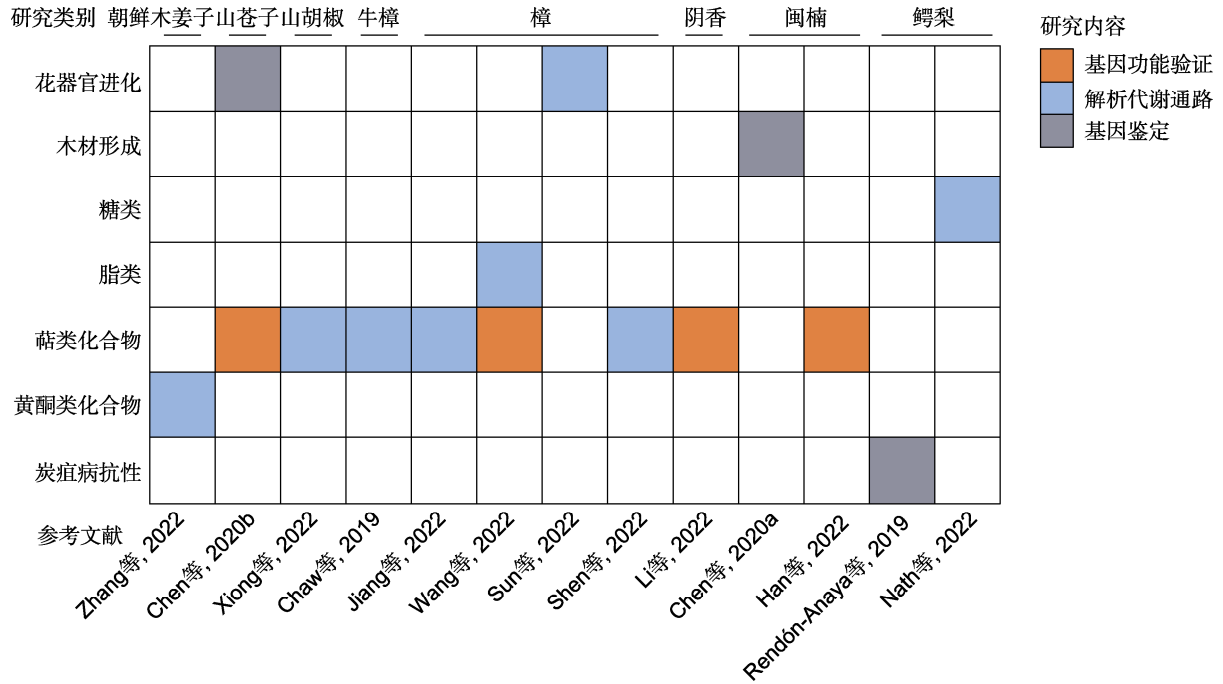


图6 樟科基因组功能基因研究热点

Figure 6 Research hotspots of functional genes of the Lauraceae genomes

个 *TPS-a* 基因。经异源转化烟草, 发现这些基因参与  $\beta$ -石竹烯、 $\alpha$ -萜烯和  $\alpha$ -古巴烯等萜类物质的合成。 $\beta$ -石竹烯是闽楠茎木质部中含量最多的萜类物质, 对3种分布最广的树木溃疡病原菌均具有抗性, 使得闽楠木材具有耐腐特性(Han et al., 2022)。研究还鉴定到 $\beta$ -石竹烯的主要合成酶PbTPS-a21和PbTPS-a25, 这对优质高抗林木种质培育具有重要应用价值(Han et al., 2022)。

作为樟科重要的基因家族, 厘清 *TPS* 基因家族在樟科中的演化历史有助于我们理解樟科的适应性进化。樟科植物基因组有6个 *TPS* 基因亚家族(*TPS-a*、*TPS-b*、*TPS-c*、*TPS-e/f*和*TPS-g*)。 *TPS-a*和*TPS-b*亚家族多样化程度最高(附录4) (Han et al., 2022; Jiang et al., 2022; Shen et al., 2022; Xiong et al., 2022)。樟科 *TPS-e/f*亚家族基因数量较其它双子叶植物显著增多, 表明樟科可能具有特异的 *TPS* 基因家族分支(Chaw et al., 2019; Xiong et al., 2022)。

串联重复、片段重复和WGD扩张的基因家族为进化提供了原始的遗传物质(Shen et al., 2022; Wang et al., 2022)。樟科物种中2次WGD事件可能造成其最近共同祖先中 *TPS* 基因大量扩张, 随后的串联重复显著促进了 *TPS* 基因家族的多样化, 特别是 *TPS-a*和*TPS-b*亚家族(Chaw et al., 2019; Chen et al., 2020b; Han et al., 2022; Jiang et al., 2022; Li et al., 2022; Wang et al., 2022)。串联重复基因表现出较高的进化速率、较近的进化历史且在染色体上分布更聚集(Jiang et al., 2022)。在闽楠基因组中已经检测到 *TPS-a*和*TPS-b*分支受到正向选择(Han et al., 2022), 从牛樟基因组中也检测到樟科特有的 *TPS-f-I*和-II子分支受到正向选择(Chaw et al., 2019)。 *TPS* 基因亚家族之间不同的进化机制促进了萜类化合物在樟科的多样化。樟科祖先物种 *TPS* 基因大量扩张, 随后不同类群的基因经历了不同的扩张和收缩过程(Han et al., 2022)。牛樟和山苍子中单萜类萜烯合酶 *TPS-b*亚家族成员较多, 而倍半萜类萜烯合酶 *TPS-a*亚家族成员较少, 但闽楠基因组中 *TPS* 基因家族扩张后保留了更多的倍半萜类合成酶 *TPS-a*亚家族成员, 丢失了大量的单萜合酶 *TPS-b*拷贝(Han et al., 2022)。 *TPS* 基因亚家族在樟科各类群中分布的差异与物种的化学组分一致, 这也揭示了樟科各分支在适应性进化过程中经历了不同的进化历史。

## 2 研究展望

已往的樟科基因组研究探讨了被子植物木兰分支的系统发生关系和花器官进化, 重建了樟科系统发生树, 阐明了次生代谢产物合成的分子机制, 取得了重要进展, 为揭示樟科的进化历史和挖掘重要性状调控基因提供了基础信息。然而, 作为重要的林木类群, 樟科基因组研究仍有许多值得进一步探讨的问题。

总体来说, 樟科基因组研究呈现3个明显的特点。(1) 樟科基因组特征信息十分匮乏。除樟族、月桂族和鳄梨族以外的樟科物种基因组大小未见报道(图2; 表1)。(2) 已测序物种明显地向中国分布及有经济价值的物种倾斜。首先, 测序物种主要集中在樟和闽楠等作为化学物质原料或提供优质木材的经济树种。其次, 2019年以来一直是我国学者在主导樟科基因组研究, 已发表的基因组中, 仅鳄梨基因组由国外学者报道, 其余全部为中国学者对国产种类的研究报道。此外, 已发表的樟科基因组集中在核心樟类群(core Lauraceae), 对樟科原始分支棠桂族、厚壳桂族、新樟族、檬果樟族及无根藤族等缺少研究(图2)。最后, 在空间分布上, 测序种类表现出明显的亚洲分布倾向, 作为樟科分布中心的热带非洲和南美洲物种(除鳄梨外)未有涉及, 这也与参与樟科基因组研究的中国学者居多有关。(3) 功能基因的挖掘不够深入。目前, 学者们侧重于挖掘重要的功能基因并解析其分子机制, 但樟科基因组研究还停留在基因层面, 缺少功能基因的验证及具体机制探究, 除 *TPS* 基因家族在樟科的起源与进化得到探讨外(Han et al., 2022), 对功能基因的系统研究和比较分析十分缺乏(图6)。此外, 与花器官进化、木材形成和代谢产物合成相关的功能基因是研究热点, 而樟科作为被子植物基部类群, 其经历漫长的历史变迁仍在森林中占据优势的独特生物学特性尚未得到足够的关注(图6)。

综合目前的研究, 我们认为未来樟科基因组研究有以下4个值得深入的方向。

(1) 调查基因组的基本特征。染色体数目以及基因组大小等基本生物学信息是基因组研究的基础(Kress et al., 2022)。染色体数目对于染色体级别的全基因组组装极为重要, 但是樟科基因组的基础研究还十分缺乏(附录2, 附录3)。物种基因组数据能够帮助我们选择适合的测序物种, 优化测序和生物信息学

研究方案,以回答特定的生物学问题(Kress et al., 2022)。全基因组大小的多样性不仅在塑造植物基因组的进化中起重要作用,而且在生态系统水平上影响植物群落的组合(Pellicer et al., 2018)。结合染色体数据研究基因组大小变化、染色体数量变化以及两者之间的相互作用机制有助于解释樟科多样化过程及基因组变化对其适应性进化的意义(Moeglein et al., 2020)。

(2) 增加测序基因组分支的代表性。目前樟科测序的基因组仅涉及3个族,还有6个族尚无代表物种的基因组发表(图2; 表1)。樟科具有复杂的进化历史,从系统发生拓扑结构不一致中可窥探一二。有代表性的物种取样能够帮助厘清樟科的系统发生关系。注释良好的基因组数据为使用重测序和靶向捕获的方法从标本中获取大量的基因序列提供参考,这将解决樟科采集和鉴定难题,从而构建更具有代表性的系统发生树(Sun et al., 2022a)。一方面,解析樟科系统发生关系不仅可帮助我们进化角度探讨特定功能基因的生物学特性和分子机制,还能为保护濒危植物提供理论依据。桫欏(*Alsophila spinulosa*)基因组就是1个良好的案例(Huang et al., 2022)。研究者基于9个群体107个桫欏个体的基因组重测序解析树蕨类植物树干形成发育的独特性,为理解蕨类植物的进化提供了新视角,也从影响桫欏种群变化的因素中拓展了保护工作思路。另一方面,基于系统发生树和分布数据可揭示樟科的时空分布格局,解析其类群多样分布类型的成因。例如,泛热带分布的厚壳桂属和琼楠属(*Beilschmiedia*) (Chanderbali et al., 2001), 东亚-北美间断分布的檫木属(*Sassafras*) (Yang et al., 2022b; Qin et al., 2023), 亚洲-大洋洲分布的土楠属(*Endiandra*) (Li et al., 2020), 以及仅分布在亚洲的新樟属(*Neocinnamomum*)和仅分布在美洲的桂土楠群(Chanderbali et al., 2001; Liu et al., 2021)。此外,大量遗传信息和代表性取样能够为解析樟科复杂的进化历史提供支持(Liu et al., 2021; Tian et al., 2021)。目前,樟科科内的多倍化只在山胡椒基因组中检测到(图5),考虑到月桂族多样的染色体变化及其更大的基因组(图3),增加月桂族物种全基因组测序,有助于理解多倍化在樟科进化中的作用。

(3) 关注具有特殊价值的物种,如月桂和无根藤(*Cassytha filiformis*)。月桂是唯一分布在欧洲且被赋

予重要文化内涵的樟科植物,其基因组具有复杂的染色体变化(附录2) (Chanderbali et al., 2001)。随着测序技术的进步及生物信息学软件的开发,越来越多的复杂基因组得到解析,如基因组大小为25.4 Gb、转座子含量达69.4%的油松(Niu et al., 2022); 已测序基因组最大的真双子叶植物杨山牡丹(*Paeonia ostii*),其基因组大小为12.8 Gb (Yuan et al., 2022); 以及六倍体燕麦(*Avena sativa*) (Peng et al., 2022)。在新技术的推动下,解析月桂基因组可探知樟科染色体结构的进化历史。而在原始被子植物樟科中,无根藤寄生藤本的生活型一直备受科学家关注。解析无根藤的基因组将与已发表的寄生植物寄生花(*Sapria himalayana*) (Cai et al., 2021)和原野菟丝子(*Cuscuta campestris*) (Vogel et al., 2018)等共同为探究寄生植物的进化提供新视角。

(4) 对功能基因进行深入挖掘。樟科基因组功能基因的挖掘集中在与花器官进化、木材形成和代谢产物合成相关的基因,且已挖掘的功能基因多数还未得到验证,不利于基因的后续应用(图6)。解析与关键特征起源相关的功能基因和基因家族,如落叶习性、单性花和寄生习性应受到更多关注。樟科因其特殊的系统发生位置,在研究生物学特征的起源与进化上具有重要意义,解析其关键特征起源的分子机制有助于揭开其作为原始类群至今繁盛的秘密。推动樟科重要物种高质量泛基因组研究将加深我们对其多样性的理解,并发现与重要性状和环境适应相关的基因和变异,为分子机制研究、品种选育和分子育种提供指导(Sun et al., 2022b; 郝晨路等, 2022)。以樟科重要的经济树种樟为例,已发表的4个樟基因组在杂合度、编码基因数量和重复序列比例上具有明显差异(表1),表现出个体间的遗传多样性(Jiang et al., 2022; Wang et al., 2022; Sun et al., 2022b; Shen et al., 2022)。泛基因组研究有助于我们理解种内个体间的差异,进而更好地利用樟树资源。虽然目前泛基因组研究主要集中在农作物,截至2021年末林木物种中仅苹果(*Malus*) (Wang et al., 2023b)和杨属(*Populus*) (Zhang et al., 2019)开展了泛基因组研究(郝晨路等, 2022)。但基因组规模大、复杂度高的林木物种泛基因组研究仍极具挑战性(Wang et al., 2023a)。随着技术的进步,林木物种的泛基因组研究不仅能够促进对乔木起源的认识,还能够解析森林建群种应对环境异

质性的分子机制, 这对理解森林对环境变化的响应具有重要意义(郝晨路等, 2022; Wang et al., 2023a)。

## 作者贡献声明

杨智: 分析数据并撰写论文; 杨永: 技术支持并修改论文。

## 参考文献

- Amborella Genome Project** (2013). The *Amborella* genome and the evolution of flowering plants. *Science* **342**, 1241–089.
- Beech E, Rivers MC, Oldfield S, Smith PP** (2017). Global-TreeSearch: the first complete global database of tree species and country distributions. *J Sustain For* **36**, 454–489.
- Blaxter M, Archibald JM, Childers AK, Coddington JA, Crandall KA, Di Palma F, Durbin R, Edwards SV, Graves JAM, Hackett KJ, Hall N, Jarvis ED, Johnson RN, Karlsson EK, Kress WJ, Kuraku S, Lawniczak MKN, Lindblad-Toh K, Lopez JV, Moran NA, Robinson GE, Ryder OA, Shapiro B, Soltis PS, Warnow T, Zhang GJ, Lewin HA** (2022). Why sequence all eukaryotes? *Proc Natl Acad Sci USA* **119**, e2115636118.
- Cai LM, Arnold BJ, Xi ZX, Khost DE, Patel N, Hartmann CB, Manickam S, Sasirat S, Nikolov LA, Mathews S, Sackton TB, Davis CC** (2021). Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). *Curr Biol* **31**, 1002–1011.
- Chanderbali AS, van der Werff H, Renner SS** (2001). Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes. *Ann Mo Bot Gard* **88**, 104–134.
- Chaw SM, Liu YC, Wu YW, Wang HY, Lin CYI, Wu CS, Ke HM, Chang LY, Hsu CY, Yang HT, Sudianto E, Hsu MH, Wu KP, Wang LN, Leebens-Mack JH, Tsai IJ** (2019). Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. *Nat Plants* **5**, 63–73.
- Chen SP, Sun WH, Xiong YF, Jiang YT, Liu XD, Liao XY, Zhang DY, Jiang SZ, Li Y, Liu B, Ma L, Yu X, He L, Liu B, Feng JL, Feng LZ, Wang ZW, Zou SQ, Lan SR, Liu ZJ** (2020a). The *Phoebe* genome sheds light on the evolution of magnoliids. *Hortic Res* **7**, 146.
- Chen YC, Li Z, Zhao YX, Gao M, Wang JY, Liu KW, Wang X, Wu LW, Jiao YL, Xu ZL, He WG, Zhang QY, Liang CK, Hsiao YY, Zhang DY, Lan SR, Huang LQ, Xu W, Tsai WC, Liu ZJ, Van de Peer Y, Wang YD** (2020b). The *Litsea* genome and the evolution of the laurel family. *Nat Commun* **11**, 1675.
- Christianson DW** (2017). Structural and chemical biology of terpenoid cyclases. *Chem Rev* **117**, 11570–11648.
- Chung KF, Hsieh CL** (2023). Synopsis of *Camphora* (Cinnamomeae, Lauraceae) of Taiwan, with two new combinations and one new synonym. *Taiwania* **68**, 384–390.
- Dong Y, Duan SC, Xia QJ, Liang ZC, Dong X, Margaryan K, Musayev M, Goryslavets S, Zdunić G, Bert PF, Lacombe T, Maul E, Nick P, Bitskinashvili K, Bisztray GD, Drori E, De Lorenzis G, Cunha J, Popescu CF, Arroyo-Garcia R, Arnold C, Ergül A, Zhu YF, Ma C, Wang SF, Liu SQ, Tang L, Wang CP, Li DW, Pan YB, Li JX, Yang L, Li XZ, Xiang GS, Yang ZJ, Chen BZ, Dai ZW, Wang Y, Arakelyan A, Kulihev V, Spotar G, Girollet N, Delrot S, Ollat N, This P, Marchal C, Sarah G, Laucou V, Bacilieri R, Röckel F, Guan PY, Jung A, Riemann M, Ujmajuridze L, Zakalashvili T, Maghradze D, Höhn M, Jahnke G, Kiss E, Deák T, Rahimi O, Hübner S, Grassi F, Mercati F, Sunseri F, Eiras-Dias J, Dumitru AM, Carrasco D, Rodriguez-Izquierdo A, Muñoz G, Uysal T, Özer C, Kazan K, Xu ML, Wang YY, Zhu SS, Lu J, Zhao MX, Wang L, Jiu S, Zhang Y, Sun L, Yang HM, Weiss E, Wang SP, Zhu YY, Li SH, Sheng J, Chen W** (2023). Dual domestications and origin of traits in grapevine evolution. *Science* **379**, 892–901.
- Eckardt NA, Birchler JA, Brady SM, Buell CR, Leebens-Mack JH, Meyers BC** (2021). Focus on the biology of plant genomes. *Plant Cell* **33**, 781–782.
- Editorial Committee of Flora of China, Chinese Academy of Sciences** (1982). Flora Reipublicae Popularis Sinicae, Vol. 31. Beijing: Science Press. pp.1–462. (in Chinese)
- 中国科学院中国植物志编辑委员会** (1982). 中国植物志, 第31卷. 北京: 科学出版社. pp. 1–462.
- Gao SH, Yu HY, Wu SY, Wang S, Geng JN, Luo YF, Hu SN** (2018). Advances of sequencing and assembling technologies for complex genomes. *Hereditas* **40**, 944–963. (in Chinese)
- 高胜寒, 禹海英, 吴双阳, 王森, 耿佳宁, 骆迎峰, 胡松年** (2018). 复杂基因组测序技术研究进展. *遗传* **40**, 944–963.
- Ge S** (2022). A review of recent studies of plant systematics and evolution in China. *Biodiv Sci* **30**, 22385. (in Chinese)
- 葛颂** (2022). 中国植物系统和进化生物学研究进展. *生物多样性* **30**, 22385.

- Guo C, Luo Y, Gao LM, Yi TS, Li HT, Yang JB, Li DZ (2023). Phylogenomics and the flowering plant tree of life. *J Integr Plant Biol* **65**, 299–323.
- Han X, Zhang JH, Han S, Chong SL, Meng GL, Song MY, Wang Y, Zhou SC, Liu CC, Lou LH, Lou XZ, Cheng LJ, Lin EP, Huang HH, Yang Q, Tong ZK (2022). The chromosome-scale genome of *Phoebe bournei* reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. *Plant Commun* **3**, 100410.
- Hao CL, Yu XF, Qu MH, Lai EH, Guo SM, Gao L (2022). Current status and prospects of pan-genome studies in plants. *Plant Sci J* **40**, 124–132. (in Chinese)
- 郝晨路, 於晓芬, 曲明昊, 赖恩惠, 郭素敏, 高磊 (2022). 植物泛基因组研究进展与展望. 植物科学学报 **40**, 124–132.
- Hu HY, Sun PC, Yang YZ, Ma JX, Liu JQ (2023a). Genome-scale angiosperm phylogenies based on nuclear, plastome, and mitochondrial datasets. *J Integr Plant Biol* **65**, 1479–1489.
- Hu RY, Li XD, Hu Y, Zhang RJ, Lv Q, Zhang M, Sheng XY, Zhao F, Chen ZJ, Ding YH, Yuan H, Wu XF, Xing S, Yan XY, Bao F, Wan P, Xiao LH, Wang XQ, Xiao W, Decker EL, van Gessel N, Renault H, Wiedemann G, Horst NA, Haas FB, Wilhelmsson PKI, Ullrich KK, Neumann E, Lv B, Liang CZ, Du HL, Lu HW, Gao Q, Cheng ZK, You HL, Xin PY, Chu JF, Huang CH, Liu Y, Dong SS, Zhang LS, Chen F, Deng L, Duan FZ, Zhao WJ, Li K, Li ZF, Li XR, Cui HJ, Zhang YE, Ma C, Zhu RL, Jia Y, Wang MZ, Hasebe M, Fu JZ, Goffinet B, Ma H, Rensing SA, Reski R, He YK (2023b). Adaptive evolution of the enigmatic *Takakia* now facing climate change in Tibet. *Cell* **186**, 3558–3576.
- Huang JF, Li L, Li J (2016). Polymorphism of the internal transcribed spacer of nrDNA in *Cinnamomum* Schaeffer (Lauraceae). *Chin Bull Bot* **51**, 609–619. (in Chinese)
- 黄建峰, 李朗, 李捷 (2016). 樟属植物ITS序列多态性分析. 植物学报 **51**, 609–619.
- Huang X, Wang WL, Gong T, Wickell D, Kuo LY, Zhang XT, Wen JL, Kim H, Lu FC, Zhao HS, Chen S, Li H, Wu WQ, Yu CJ, Chen S, Fan W, Chen S, Bao XQ, Li L, Zhang D, Jiang LY, Yan XJ, Liao ZY, Zhou GK, Guo YL, Ralph J, Sederoff RR, Wei HR, Zhu P, Li FW, Ming R, Li QZ (2022). The flying spider-monkey tree fern genome provides insights into fern evolution and arborescence. *Nat Plants* **8**, 500–512.
- Jia QD, Brown R, Köllner TG, Fu JY, Chen XL, Wong GKS, Gershenzon J, Peters RJ, Chen F (2022). Origin and early evolution of the plant terpene synthase family. *Proc Natl Acad Sci USA* **119**, e2100361119.
- Jiang RH, Chen XL, Liao XZ, Peng D, Han XX, Zhu CS, Wang P, Hufnagel DE, Wang L, Li KX, Li C (2022). A chromosome-level genome of the camphor tree and the underlying genetic and climatic factors for its top-geotherbalism. *Front Plant Sci* **13**, 827890.
- Jing CY, Zhang FM, Wang XH, Wang MX, Zhou L, Cai Z, Han JD, Geng MF, Yu WH, Jiao ZH, Huang L, Liu R, Zheng XM, Meng QL, Ren NN, Zhang HX, Du YS, Wang X, Qiang CG, Zou XH, Gaut BS, Ge S (2023). Multiple domestications of Asian rice. *Nat Plants* **9**, 1221–1235.
- Joshi SC, Padalia RC, Bisht DS, Mathela CS (2009). Terpenoid diversity in the leaf essential oils of Himalayan Lauraceae species. *Chem Biodivers* **6**, 1364–1373.
- Kress WJ, Soltis DE, Kersey PJ, Wegrzyn JL, Leebens-Mack JH, Gostel MR, Liu X, Soltis PS (2022). Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proc Natl Acad Sci USA* **119**, e2115640118.
- Lewin HA, Robinson GE, Kress WJ, Baker WJ, Codrington J, Crandall KA, Durbin R, Edwards SV, Forest F, Gilbert MTP, Goldstein MM, Grigoriev IV, Hackett KJ, Haussler D, Jarvis ED, Johnson WE, Patrinos A, Richards S, Castilla-Rubio JC, van Sluys MA, Soltis PS, Xu X, Yang HM, Zhang GJ (2018). Earth BioGenome Project: sequencing life for the future of life. *Proc Natl Acad Sci USA* **115**, 4325–4333.
- Li FP, Huang SL, Mei Y, Wu BQ, Hou ZW, Zhan PL, Hou ZH, Huang WJ, Zhao JL, Wang JH (2022). Genome assembly provided new insights into the *Cinnamomum burmannii* evolution and D-borneol biosynthesis differences between chemotypes. *Ind Crops Prod* **186**, 115181.
- Li HT, Luo Y, Gan L, Ma PF, Gao LM, Yang JB, Cai J, Gitzendanner MA, Fritsch PW, Zhang T, Jin JJ, Zeng CX, Wang H, Yu WB, Zhang R, van der Bank M, Olmstead RG, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Yi TS, Li DZ (2021). Plastid phylogenomic insights into relationships of all flowering plant families. *BMC Biol* **19**, 232.
- Li HW, Liu B, Davis CC, Yang Y (2020). Plastome phylogenomics, systematics, and divergence time estimation of the *Beilschmiedia* group (Lauraceae). *Mol Phylogenet Evol* **151**, 106901.
- Li J, Christophel DC, Conran JG, Li HW (2004). Phylogenetic relationships within the ‘core’ Laureae (*Litsea* complex, Lauraceae) inferred from sequences of the



- chloroplast gene *matK* and nuclear ribosomal DNA ITS regions. *Plant Syst Evol* **246**, 19–34.
- Li J, Conran JG, Christophei DC, Li ZM, Li L, Li HW (2008). Phylogenetic relationships of the *Litsea* complex and core Laureae (Lauraceae) using ITS and ETS sequences and morphology. *Ann Mo Bot Gard* **95**, 580–599.
- Li J, Li XW (2004). Systematic research on the world scale advances in Lauraceae. *Acta Bot Yunnan* **26**, 1–11. (in Chinese)
- 李捷, 李锡文 (2004). 世界樟科植物系统学研究进展. 云南植物研究 **26**, 1–11.
- Liu Y, Wang SB, Li LZ, Yang T, Dong SS, Wei T, Wu SD, Liu YB, Gong YQ, Feng XY, Ma JC, Chang GX, Huang JL, Yang Y, Wang HL, Liu M, Xu Y, Liang HP, Yu J, Cai YQ, Zhang ZW, Fan YN, Mu WX, Sahu SK, Liu SC, Lang XA, Yang LL, Li N, Habib S, Yang YQ, Lindstrom AJ, Liang P, Goffinet B, Zaman S, Wegrzyn JL, Li DX, Liu J, Cui J, Sonnenschein EC, Wang XB, Ruan J, Xue JY, Shao ZQ, Song C, Fan GY, Li Z, Zhang LS, Liu JQ, Liu ZJ, Jiao YN, Wang XQ, Wu H, Wang ET, Lisby M, Yang HM, Wang J, Liu X, Xu X, Li N, Soltis PS, Van de Peer Y, Soltis DE, Gong X, Liu H, Zhang SZ (2022). The *Cycas* genome and the early evolution of seed plants. *Nat Plants* **8**, 389–401.
- Liu ZF, Ma H, Ci XQ, Li L, Song Y, Liu B, Li HW, Wang SL, Qu XJ, Hu JL, Zhang XY, Conran JG, Twyford AD, Yang JB, Hollingsworth PM, Li J (2021). Can plastid genome sequencing be used for species identification in Lauraceae? *Bot J Linn Soc* **197**, 1–14.
- Moeglein MK, Chatelet DS, Donoghue MJ, Edwards EJ (2020). Evolutionary dynamics of genome size in a radiation of woody plants. *Am J Bot* **107**, 1527–1541.
- Nagegowda DA, Gupta P (2020). Advances in biosynthesis, regulation, and metabolic engineering of plant specialized terpenoids. *Plant Sci* **294**, 110457.
- Nath O, Fletcher SJ, Hayward A, Shaw LM, Masouleh AK, Furtado A, Henry RJ, Mitter N (2022). A haplotype resolved chromosomal level avocado genome allows analysis of novel avocado genes. *Hortic Res* **9**, uhac157.
- Niu SH, Li J, Bo WH, Yang WF, Zuccolo A, Giacomello S, Chen X, Han FX, Yang JH, Song YT, Nie YM, Zhou B, Wang PY, Zuo Q, Zhang H, Ma JJ, Wang J, Wang L, Zhu QY, Zhao HH, Liu ZM, Zhang XM, Liu T, Pei SR, Li ZM, Hu Y, Yang YH, Li WZ, Zan YJ, Zhou LH, Lin JX, Yuan TQ, Li W, Li Y, Wei HR, Wu HX (2022). The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell* **185**, 204–217.
- Oginuma K, Tobe H (2006). Chromosome evolution in the Laurales based on analyses of original and published data. *J Plant Res* **119**, 309–320.
- One Thousand Plant Transcriptomes Initiative (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685.
- Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ (2018). Genome size diversity and its impact on the evolution of land plants. *Genes (Basel)* **9**, 88.
- Peng YY, Yan HH, Guo LC, Deng C, Wang CL, Wang YB, Kang LP, Zhou PP, Yu KQ, Dong XL, Liu XM, Sun ZY, Peng Y, Zhao J, Deng D, Xu YH, Li Y, Jiang QT, Li Y, Wei LM, Wang JR, Ma J, Hao M, Li W, Kang HY, Peng ZS, Liu DC, Jia JZ, Zheng YL, Ma T, Wei YM, Lu F, Ren CZ (2022). Reference genome assemblies reveal the origin and evolution of allohexaploid oat. *Nat Genet* **54**, 1248–1258.
- Pichersky E, Raguso RA (2018). Why do plants produce so many terpenoid compounds? *New Phytol* **220**, 692–702.
- Qin SY, Zuo ZY, Guo C, Du XY, Liu SY, Yu XQ, Xiang XG, Rong J, Liu B, Liu ZF, Ma PF, Li DZ (2023). Phylogenomic insights into the origin and evolutionary history of evergreen broadleaved forests in East Asia under Cenozoic climate change. *Mol Ecol* **32**, 2850–2868.
- Rendón-Anaya M, Ibarra-Laclette E, Méndez-Bravo A, Lan TY, Zheng CF, Carretero-Paulet L, Perez-Torres CA, Chacón-López A, Hernandez-Guzmán G, Chang TH, Farr KM, Barbazuk WB, Chamala S, Mutwil M, Shihhare D, Alvarez-Ponce D, Mitter N, Hayward A, Fletcher S, Rozas J, Sánchez Gracia A, Kuhn D, Barrientos-Priego AF, Salojärvi J, Librado P, Sankoff D, Herrera-Estrella A, Albert VA, Herrera-Estrella L (2019). The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. *Proc Natl Acad Sci USA* **116**, 17081–17089.
- Rice A, Glick L, Abadi S, Einhorn M, Kopelman NM, Salman-Minkov A, Mayzel J, Chay O, Mayrose I (2015). The chromosome counts database (CCDB)—a community resource of plant chromosome numbers. *New Phytol* **206**, 19–26.
- Rohwer JG (1993). Lauraceae. In: Kubitzki K, Rohwer JG, Bittrich V, eds. Flowering Plants · Dicotyledons: Magnoliid, Hamamelid and Caryophyllid Families. Berlin: Springer. pp. 366–391.
- Rohwer JG (2000). Toward a phylogenetic classification of the Lauraceae: evidence from *matK* sequences. *Syst Bot*

- 25, 60–71.
- Rohwer JG, Rudolph B (2005). Jumping genera: the phylogenetic positions of *Cassytha*, *Hypodaphnis*, and *Neocinnamomum* (Lauraceae) based on different analyses of *trnK* intron sequences. *Ann Mo Bot Gard* **92**, 153–178.
- Shen TF, Qi HR, Luan XY, Xu WL, Yu FX, Zhong YD, Xu M (2022). The chromosome-level genome sequence of the camphor tree provides insights into Lauraceae evolution and terpene biosynthesis. *Plant Biotechnol J* **20**, 244–246.
- Steenwyk JL, Li YN, Zhou XF, Shen XX, Rokas A (2023). Incongruence in the phylogenomics era. *Nat Rev Genet* **24**, 834–850.
- Sun WH, Xiang S, Zhang QG, Xiao L, Zhang DY, Zhang PL, Chen DQ, Hao Y, Liu DK, Ding L, Li YF, Ni H, Wang YF, Wu X, Liu FH, Chen GR, Han GY, Chen JZ, Su BC, Gao JX, Wan XH, Wang ZW, Chen YC, Wang YD, Huang W, Liu BB, Zou XX, Ni L, Liu ZJ, Zou SQ (2022a). The camphor tree genome enhances the understanding of magnoliid evolution. *J Genet Genomics* **49**, 249–253.
- Sun YQ, Shang LG, Zhu QH, Fan LJ, Guo LB (2022b). Twenty years of plant genome sequencing: achievements and challenges. *Trends Plant Sci* **27**, 391–401.
- The Angiosperm Phylogeny Group (2016). An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. *Bot J Linn Soc* **181**, 1–20.
- The *Arabidopsis* Genome Initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815.
- Tian YJ, Zhou JB, Zhang YY, Wang S, Wang Y, Liu H, Wang ZS (2021). Research progress in plant molecular systematics of Lauraceae. *Biology (Basel)* **10**, 391.
- Trofimov D, Cadar D, Schmidt-Chanasit J, Rodrigues de Moraes PL, Rohwer JG (2022). A comparative analysis of complete chloroplast genomes of seven *Ocotea* species (Lauraceae) confirms low sequence divergence within the *Ocotea* complex. *Sci Rep* **12**, 1120.
- Vogel A, Schwacke R, Denton AK, Usadel B, Hollmann J, Fischer K, Bolger A, Schmidt MHW, Bolger ME, Gundlach H, Mayer KFX, Weiss-Schneeweiss H, Temsch EM, Krause K (2018). Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*. *Nat Commun* **9**, 2515.
- Wakabayashi K (2000). Changes in cell wall polysaccharides during fruit ripening. *J Plant Res* **113**, 231–237.
- Wan T, Liu ZM, Leitch IJ, Xin HP, Maggs-Kölling G, Gong YB, Li Z, Marais E, Liao YY, Dai C, Liu F, Wu QJ, Song C, Zhou YD, Huang WC, Jiang K, Wang Q, Yang Y, Zhong ZX, Yang M, Yan X, Hu GW, Hou C, Su YJ, Feng SX, Yang J, Yan JJ, Chu JF, Chen F, Ran JH, Wang XQ, Van de Peer Y, Leitch AR, Wang QF (2021). The *Welwitschia* genome reveals a unique biology underpinning extreme longevity in deserts. *Nat Commun* **12**, 4247.
- Wang S, Qian YQ, Zhao RP, Chen LL, Song JM (2023a). Graph-based pan-genomes: increased opportunities in plant genomics. *J Exp Bot* **74**, 24–39.
- Wang T, Duan S, Xu C, Wang Y, Zhang X, Xu X, Chen L, Han Z, Wu T (2023b). Pan-genome analysis of 13 *Malus* accessions reveals structural and sequence variations associated with fruit traits. *Nat Commun* **14**, 7377.
- Wang W, Liu Y (2020). The current status, problems, and policy suggestions for reconstructing the plant tree of life. *Biodiv Sci* **28**, 176–188. (in Chinese)
- 王伟, 刘阳 (2020). 植物生命之树重建的现状、问题和对策建议. *生物多样性* **28**, 176–188.
- Wang XD, Xu CY, Zheng YJ, Wu YF, Zhang YT, Zhang T, Xiong ZY, Yang HK, Li J, Fu C, Qiu FY, Dai XY, Liu XL, He XS, Zhou SS, Li SX, Fu T, Xie H, Chen YL, Zhang QQ, Wang HQ, Wang YD, Zhou C, Jiang XM (2022). Chromosome-level genome assembly and resequencing of camphor tree (*Cinnamomum camphora*) provides insight into phylogeny and diversification of terpenoid and triglyceride biosynthesis of *Cinnamomum*. *Hortic Res* **9**, uhac216.
- Weigelt P, König C, Kreft H (2020). GIFT—a global inventory of floras and traits for macroecology and biogeography. *J Biogeogr* **47**, 16–43.
- Wu S, Sun HH, Gao L, Branham S, McGregor C, Renner SS, Xu Y, Kousik C, Wechter WP, Levi A, Fei Z (2023). A *Citrullus* genus super-pangenome reveals extensive variations in wild and cultivated watermelons and sheds light on watermelon evolution and domestication. *Plant Biotechnol J* **21**, 1926–1928.
- Wu SD, Han BC, Jiao YN (2020). Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Mol Plant* **13**, 59–71.
- Xiao TW, Ge XJ (2022). Plastome structure, phylogenomics, and divergence times of tribe Cinnamomeae (Lauraceae). *BMC Genomics* **23**, 642.
- Xiao TW, Xu Y, Jin L, Liu TJ, Yan HF, Ge XJ (2020). Conflicting phylogenetic signals in plastomes of the tribe Laureae (Lauraceae). *PeerJ* **8**, e10155.

- Xiao TW, Yan HF, Ge XJ** (2022). Plastid phylogenomics of tribe Perseeae (Lauraceae) yields insights into the evolution of East Asian subtropical evergreen broad-leaved forests. *BMC Plant Biol* **22**, 32.
- Xie LJ, Ye CY, Shen EH** (2021). Advances in plant genome construction. *Plant Sci J* **39**, 681–691. (in Chinese)
- 谢玲娟, 叶楚玉, 沈恩惠 (2021). 植物基因组测序研究进展. 植物科学学报 **39**, 681–691.
- Xiong B, Zhang LM, Xie L, Li LZ, He XX, Niu Y, Zhang TY, Liao S, Dong SB, Zhang ZX** (2022). Genome of *Lindera glauca* provides insights into the evolution of biosynthesis genes for aromatic compounds. *iScience* **25**, 104761.
- Yang Z, Liu B, Yang Y, Ferguson DK** (2022a). Phylogeny and taxonomy of *Cinnamomum* (Lauraceae). *Ecol Evol* **12**, e9378.
- Yang Z, Tan C, Wei YM, Rohwer JG, Liu B, Yang Y** (2022b). Floral morphology and phenology of *Sassafras tzumu* (Lauraceae). *BMC Plant Biol* **22**, 327.
- Yuan JH, Jiang SJ, Jian JB, Liu MY, Yue Z, Xu JB, Li J, Xu CY, Lin LH, Jing Y, Zhang XX, Chen HX, Zhang LJ, Fu T, Yu SY, Wu ZY, Zhang Y, Wang CZ, Zhang X, Huang LB, Wang HQ, Hong DY, Chen XY, Hu YH** (2022). Genomic basis of the giga-chromosomes and giga-genome of tree peony *Paeonia ostii*. *Nat Commun* **13**, 7328.
- Zhang BH, Yao XZ, Chen HF, Lu LT** (2022). High-quality chromosome-level genome assembly of *Litsea coreana* L. provides insights into Magnoliids evolution and flavonoid biosynthesis. *Genomics* **114**, 110394.
- Zhang BY, Zhu WX, Diao S, Wu XJ, Lu JQ, Ding CJ, Su XH** (2019). The poplar pangenome provides insights into the evolutionary history of the genus. *Commun Biol* **2**, 215.
- Zhang XT, Wang G, Zhang SC, Chen S, Wang YB, Wen P, Ma XK, Shi Y, Qi R, Yang Y, Liao ZY, Lin J, Lin JS, Xu XM, Chen XQ, Xu XD, Deng F, Zhao LH, Lee YL, Wang R, Chen XY, Lin YR, Zhang JS, Tang HB, Chen J, Ming R** (2020). Genomes of the banyan tree and pollinator wasp provide insights into fig-wasp coevolution. *Cell* **183**, 875–889.

## Research Advances on Nuclear Genomes of Economically Important Trees of Lauraceae

Zhi Yang<sup>1, 2, 3</sup>, Yong Yang<sup>1, 2, 3\*</sup>

<sup>1</sup>Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China

<sup>2</sup>College of Biology and the Environment, Nanjing Forestry University, Nanjing 210037, China; <sup>3</sup>Key Laboratory of State Forestry and Grassland Administration on Subtropical Forest Biodiversity Conservation, Nanjing Forestry University, Nanjing 210037, China

**Abstract** With the rapid development of sequencing technology and bioinformatic analysis tools, the studies of plant genomes has been growing in an unprecedented speed. The family Lauraceae belongs to Magnoliids and is mainly found in pantropics with high species diversity. Species of this family are mostly woody and economically and ecologically important. Up to now, 13 genomes belonging to 8 species in the family Lauraceae have been published. Here, we review the recent advances on the published nuclear genomes of Lauraceae by focusing on their characteristics, evolution and phylogeny, and the functional genes and gene families. We highlight the recent achievements in the whole genome duplication events, and the genes/gene families related to the floral organ determination and metabolites in Lauraceae. Finally, we suggest further studies should be performed to improve our understanding of the function and evolution of this family by sequencing more species as well as more representative samples in all major clades, especially by tackling species with special value, and studying the species-specific functional genes.

**Key words** Lauraceae, evolution, nuclear genome, function gene, phylogeny

**Yang Z, Yang Y (2024).** Research advances on nuclear genomes of economically important trees of Lauraceae. *Chin Bull Bot* 59, 302–318.

\* Author for correspondence. E-mail: yongyang@njfu.edu.cn

(责任编辑: 白羽红)

**附录 1** 已测序维管植物科的生活型数量分布

**Appendix 1** The distribution of the number of growth form in the families of sequenced vascular plants

**附录 2** 樟科物种染色体数目信息

**Appendix 2** Chromosome number of Lauraceae species

**附录 3** 流式细胞仪测定的樟科基因组大小

**Appendix 3** Genome size of Lauraceae estimated by flow cytometry

**附录 4** 樟科 TPS 基因组亚家族数量

**Appendix 4** Number of subfamilies in Lauraceae TPS genomes

<https://www.chinbullbotany.com/fileup/1674-3466/PDF/23-035-1.pdf>

### 通讯作者/团队简介

杨永, 南京林业大学教授, 世界自然保护联盟(IUCN)物种生存委员会松柏类专家组主席, 博士生导师。长期从事植物分类、生物地理和濒危物种保护研究。发表论文(著) 150 余篇(部), 主持国家自然科学基金和江苏省自然科学基金等项目。主编《世界裸子植物的分类和地理分布》获第十五届“上海图书奖”一等奖。目前其研究团队以裸子植物和樟科为研究对象, 开展整合分类学、生物地理和生物多样性保护等方面研究。