

# 被子植物系统发育深层关系研究: 进展与挑战

曾丽萍 张 宁 马 红\*

(复旦大学生命科学院, 遗传工程国家重点实验室, 植物科学研究所, 进化生物学中心, 上海 200433)

**摘要:** 被子植物系统发育学是研究被子植物及其各类群间亲缘关系与进化历史的学科。从20世纪90年代起, 核苷酸和氨基酸序列等分子数据开始被广泛运用于被子植物系统发育研究, 经过20多年的发展, 从使用单个或联合少数几个细胞器基因, 到近期应用整个叶绿体基因组来重建被子植物的系统发育关系, 目、科水平上的被子植物系统发育框架已被广泛接受。在这个框架中, 基部类群、主要的5个分支(即真双子叶植物、单子叶植物、木兰类、金粟兰目和金鱼藻目)、每个分支所包含的目以及几个大分支包括的核心类群等都具有高度支持。与此同时, 细胞器基因还存在一些固有的问题, 例如单亲遗传、系统发育信息量有限等, 因此近年来双亲遗传的核基因在被子植物系统发育研究中的重要性逐渐得到关注, 并在不同分类阶元的研究中都取得了一定进展。但是, 被子植物系统发育中仍然存在一些难以确定的关系, 例如被子植物5个分支之间的关系、真双子叶植物内部某些类群的位置等。本文简述了20多年来被子植物系统发育深层关系的主要研究进展, 讨论了被子植物系统发育学常用的细胞器基因和核基因的选用, 已经确定和尚未确定系统发育位置的主要类群, 以及研究中尚存在的问题和可能的解决方法。

**关键词:** 亲缘与进化关系, 分子系统发育, 细胞器基因, 核基因, 系统发育基因组学

## Advances and challenges in resolving the angiosperm phylogeny

Liping Zeng, Ning Zhang, Hong Ma\*

*State Key Laboratory of Genetic Engineering and Institute of Plant Biology, Center for Evolutionary Biology, School of Life Sciences, Fudan University, Shanghai 200433*

**Abstract:** Angiosperm phylogenetics investigates the evolutionary history and relationships of angiosperms based on the construction of phylogenetic trees. Since the 1990s, nucleotide or amino acid sequences have been widely used for this and angiosperm phylogenetic analysis has advanced from using single or a combination of a few organellar genes to whole plastid genome sequences, resulting in the widely accepted modern molecular systematics of angiosperms. The current framework of the angiosperm phylogeny includes highly supported basal angiosperm relationships, five major clades (eudicots, monocots, magnoliids, Chloranthales, and Ceratophyllales), orders grouped within these clades, and core groups in the monocots or eudicots. However, organellar genes have some limitations; these involve uniparental inheritance in most instances and a relatively low percentage of phylogenetic informative sites. Thus, they are unable to resolve some relationships even when whole plastid genome sequences are used. Therefore, the utility of biparentally inherited nuclear genes with more information about evolutionary history, has gradually received more attention. Nevertheless, there are still some plant groups that are difficult to place in the angiosperm phylogeny, such as those involving the relative positions of the five major groups as well as those of several orders of eudicots. In this review, we discuss the applications, advantages and disadvantages of marker genes, the deep relationships that have been resolved in angiosperm phylogeny, groups with uncertain positions, and the challenges that remain in resolving an accurate phylogeny for angiosperms.

**Key words:** evolutionary relationships, molecular phylogeny, organellar genes, nuclear genes, phylogenomics

收稿日期: 2013-08-16; 接受日期: 2013-12-07

基金项目: 植物适应逆境中转录调控的分子进化机制(91131007)。

\* 通讯作者 Author for correspondence. E-mail: hongma@fudan.edu.cn; hxm16@psu.edu

被子植物也称有花植物,约起源于一亿八千万年前的早侏罗纪(early Jurassic)甚至可能更早(Sun *et al.*, 2002; Crane *et al.*, 2004; Smith *et al.*, 2010),并在短时间内快速分化形成种类繁多、形态各异的类群(Davies *et al.*, 2004; Magallón & Castillo, 2009)。早期被子植物的过渡型化石很少被发现,因此,被子植物在演化过程中呈现出的快速起源及其在较短地质时期内物种多样化的现象被达尔文称为“令人烦恼之谜”(abominable mystery)(Darwin & Seward, 1903; Friedman, 2009)。已知的被子植物约有35万种,占整个植物界的一半,在现今陆地生态系统中占绝对优势,为人类提供氧气、食物、木材、纤维、医药和能源等重要生存物资(Tilman *et al.*, 2002),也为其他生物类群如昆虫(Farrell, 1998; Wilf *et al.*, 2000; Moreau *et al.*, 2006)、两栖类(Roelants *et al.*, 2007)、哺乳动物(Bininda-Emonds *et al.*, 2007; Roelants *et al.*, 2007)和蕨类植物(Schneider *et al.*, 2004)的扩张或再次繁荣带来新的契机。

被子植物在生长习性、形态特征、解剖构造、生理性状以及生殖方式等方面均呈现极大的多样性和适应性(Crane & Lidgard, 1989; Endress & Doyle, 2009)。一个能够真实反映被子植物间亲缘关系的系统发育树对于了解被子植物的起源、物种扩张及性状的起源与演化历史(如花的起源与演化)等都至关重要。同时,被子植物的繁盛与现代陆地生态系统密切相关,可靠的被子植物系统发育树也是其他生命科学研究的基础,可为作物改良、疾病预防、生物多样性保护等提供重要的理论依据。

由于受研究手段的限制,早期分类学家主要依赖于形态学、细胞学和生物化学等方面的证据进行分类,但是形态性状有很强的环境可塑性,以至于趋同进化(convergent evolution)和平行进化(parallel evolution)多次发生,在一定程度上影响了它们的适用范围;而且不同研究者判断形态性状相似性的标准有差异,因此存在很多分类关系有争议的物种或类群。分子生物学和计算科学的快速发展,使得利用核苷酸或氨基酸等分子性状阐明被子植物间的亲缘关系成为可能。经过二十余年的发展,分子系统学研究也从最初的基于单个基因(Hamby & Zimmer, 1992; Chase *et al.*, 1993; Hilu & Liang, 1997)向联合多个基因(Qiu *et al.*, 1999; Soltis *et al.*, 1999),乃至整个细胞器基因组的方向发展(Jansen

*et al.*, 2007; Moore *et al.*, 2007, 2010)。世界各地的植物系统学家通力合作,构建了现代被子植物系统发育框架(Judd & Olmstead, 2004; Soltis *et al.*, 2005, 2011; Soltis & Soltis, 2013),例如提出并不断改善了被子植物APG(1998, 2003, 2009)分类系统,并已更新至APG III(APG, 1998, 2003; Bremer *et al.*, 2009),使人们对被子植物主要类群间的亲缘关系以及演化过程有了全新的认识。

本文综述了被子植物系统发育研究常用的标记基因、被子植物系统关系研究的现状及研究中存在的问题和可能的解决办法等,以期为国内关注被子植物系统发育关系研究的学者提供一定的参考。

## 1 被子植物系统发育研究的标记基因

### 1.1 细胞器基因和其基因间隔区

被子植物系统发育研究早期,细胞器(叶绿体、线粒体)基因或其基因间隔区因为较易获得而被广泛使用。

叶绿体是植物特有的细胞器,其环状基因组DNA被两个反向重复序列(inverted repeat sequence, IR)分为大单拷贝区(large single copy, LSC)和小单拷贝区(small single copy, SSC)。虽然不同物种中叶绿体DNA(cpDNA)拷贝数目有差异,但基因组成和排列相似,基因数目几乎相同,不同物种中的叶绿体基因是直系同源基因(ortholog),因此利用叶绿体基因构建系统发育树时很少像使用核基因那样受旁系同源基因(paralog)的干扰;同时,叶绿体基因序列相对比较保守,便于扩增和克隆,因此在被子植物系统发育研究中得到广泛应用(Olmstead & Palmer, 1994)。例如Chase等(1993)使用叶绿体基因*rbcL*分析了代表种子植物各类群的499个物种,全面探讨了被子植物主要类群的系统发育关系,是植物系统发育研究的典范。在叶绿体基因中,*rbcL*, *atpB*, *ndhF*和*matK*是比较常用的被子植物系统发育标记基因(Kim & Jansen, 1995; Soltis *et al.*, 2000; Hilu *et al.*, 2003; Qiu *et al.*, 2005; Burleigh *et al.*, 2009),其中*atpB*和*rbcL*长度接近(氨基酸编码区序列长度大约为1,500 bp),其核苷酸替换速率(substitution site/unit time)约为0.068–0.108,序列易于比对(alignment);*ndhF*的序列更长(> 2,000 bp),进化速率也相对较快;*matK*进化速率在叶绿体基因中最高(≥1,500 bp, ~0.141),更适用于研究相对近缘物种

之间的系统关系,但因其插入/缺失(indel)比较多,较其他叶绿体基因更难比对。线粒体基因如*mtSSU*, *mtLSU*, *atpA*, *cox1*, *matR*, *atp1*, *nad5*和*rps3*等也被经常使用(Qiu *et al.*, 1999; Meng *et al.*, 2002; Zanis *et al.*, 2002),其中*atp1*, *rps3*, *nad5*和*matR*等的进化速率比较慢,核酸替换速率约为0.029–0.044,对于研究远缘物种之间的亲缘关系很有帮助(Qiu *et al.*, 2010)。

然而,基于不同单基因构建的被子植物系统发育树之间往往存在分歧,而且很多分支的支持率并不高,这主要是由于单基因的序列较短和信息位点太少,造成了随机误差(stochastic error)(Rokas *et al.*, 2003; Delsuc *et al.*, 2005; Jeffroy *et al.*, 2006)。例如Savolainen等(2000)使用*atpB*序列构建被子植物系统发育树时认为无油樟(*Amborella trichocarpa*)和睡莲目(Nymphaeales)是被子植物最基部的类群,而Chase等(1993)基于*rbcL*序列的结果则显示金鱼藻(*Ceratophyllum*)位于被子植物最基部。随着数据的积累,植物系统学家们联合几个单基因的序列来重建被子植物系统发育关系,发现联合基因的系统树分支间的关系更明确,各分支的支持率也得到了显著提高(Qiu *et al.*, 1999; Soltis *et al.*, 1999, 2002; Fishbein *et al.*, 2001)。

测序技术的成熟与计算能力的提高促使越来越多的植物叶绿体基因组被测序,叶绿体系统发育基因组学(phylogenomics)也逐步成为被子植物系统发育研究的趋势之一(Moore *et al.*, 2010; Jansen *et al.*, 2011; Weng *et al.*, 2012)。例如, Jansen等(2007)利用64个物种的叶绿体基因组的81个基因重建了被子植物大类群之间的亲缘关系; Moore等(2010)根据86个物种的83个叶绿体基因重建了真双子叶(eudicots)植物内部大类群之间的亲缘关系。叶绿体基因组也开始被应用于被子植物科以下分类阶元的进化研究。例如Zhang等(2011)成功运用叶绿体基因组序列研究了竹亚科(Bambusoideae)在禾本科(Poaceae)中的系统位置,提出了竹亚科与早熟禾亚科(Pooideae)为姊妹群的假说; Wu和Ge(2012)利用22个禾本科植物的叶绿体基因组成功解决了禾本科三大类群即竹亚科、早熟禾亚科和稻亚科(Ehrhartoidea)之间的亲缘关系; Xi等(2012)对约占热带雨林物种数40%的金虎尾目(Malpighiales)进行了系统发育重建,并鉴定出了新的12个分支; Sun等(2013)推断基部真双子叶植物昆栏树科(Trochon-

dendraceae)在30–44百万年前分化形成昆栏树属(*Trochodendron*)和水青树属(*Tetracentron*)。

线粒体包含的遗传物质应用于植物系统发育研究中,能从另一个细胞器的遗传角度来揭示被子植物的系统发育关系。例如Zhu等(2007)利用线粒体基因*matR*研究蔷薇类各分支的关系,认为COM分支与锦葵类为姊妹群,桃金娘目(Myrtales)与牻牛儿苗目(Geraniales)依次位于其他蔷薇类的基部,而非形成姊妹群; Qiu等(2010)应用4个线粒体基因(*atp1*, *matR*, *nad5*和*rps3*)也获得了同样的结果,同时上述结果也得到了核基因的支持。但是相对于叶绿体系统发育基因组学的发展,植物线粒体基因组的一些特性限制了其在系统发育研究中的应用(Palmer, 1988)。例如线粒体基因组在不同植物类群中的大小变异很大(300–600 kb),且含有很多较短的同源序列,这些同源序列之间会通过分子内重组(intramolecular recombination)而形成小的亚基因组环状DNA,与完整的线粒体基因组共存于细胞中(Lonsdale *et al.*, 1988)。虽然线粒体基因组结构变异很大,但是植物线粒体基因的核苷酸替换率比叶绿体基因组慢4倍(Palmer & Herbon, 1988);而且,在部分寄生类群中存在基因水平转移现象(horizontal gene transfer),这使得线粒体基因组中有核基因的插入(Bergthorsson *et al.*, 2003; Davis & Wurdack, 2004; Westwood *et al.*, 2010)。因此,选取合适的线粒体基因作为标记基因是植物系统发育研究中需要仔细考虑的问题。

## 1.2 核糖体RNA基因和其基因间隔区及编码蛋白的核基因

虽然细胞器基因,尤其是叶绿体基因组在研究被子植物不同分类阶元的物种亲缘关系中取得了极为重要的进展,但是近年来对于仅使用细胞器基因研究被子植物间的亲缘关系开始产生争议。在大多数被子植物类群中,细胞器基因是单亲遗传的,叶绿体为双亲遗传的物种仅占约14%(Corriveau & Coleman, 1988)。单亲遗传的细胞器基因只能反映母系或父系的进化历史(Day & Ellis, 1984; Fauré *et al.*, 1994),不能完全揭示被子植物的进化历史;在寄生型的被子植物中,叶绿体基因可能会部分丢失(Palmer, 1990),线粒体基因也存在基因水平转移现象(Bergthorsson *et al.*, 2003; Won & Renner, 2003; Davis & Wurdack, 2004);而且,细胞器基因系统发

育信息有限,即使应用了整个叶绿体基因组数据,被子植物某些关键类群的系统位置仍然无法确定(Moore *et al.*, 2010),因此基于双亲遗传且蕴含丰富系统发育信息的核基因的系统发育研究结果显得越来越重要(Sang, 2002; Lee *et al.*, 2011; Zimmer & Wen, 2013)。另外,核基因的使用也有助于发现和分析被子植物进化过程中的杂交(hybridization)、多倍化(polyploidization)和渐渗(introgression)等现象,并且便于将被子植物系统发育的研究结果整合到整个生物界从而构建出生命之树(tree of life)。

被子植物系统发育研究早期,核糖体DNA(rDNA)(Hamby & Zimmer, 1992)和其内转录间隔区(internal transcribed spacer, ITS)(Baldwin *et al.*, 1995)是常用的分子标记。例如Hamby和Zimmer(1992)最早基于18S和26S rDNA基因片段构建了包含60个物种的被子植物系统发育树,认为原来基于形态学定义的双子叶植物并不是单系群(monophyletic group),并提出了真双子叶植物的概念。rDNA编码区进化速率慢,例如18S rDNA在植物中的碱基替换速率约为0.077(Qiu *et al.*, 2010),适用于目或科之间分类阶元的系统发育重建,但在鉴定亲缘关系近的种属关系时则因太保守而不适用;ITS的进化速率很快,在亲缘关系远的物种中因序列差异太大而难以比对(Álvarez & Wendel, 2003),但适用于解决科内或属内分类阶元之间的亲缘关系,而且是很好的DNA条形码(barcoding)分子标记(Li *et al.*, 2011)。然而,由于在某些物种中协同进化(concerted evolution)不完全,导致同一个体中不同ITS序列之间出现分化,必须依靠克隆才能完成测序,不但增加了工作量,而且使选择类群的代表序列成为难题(Álvarez & Wendel, 2003)。基于上述原因,寻找更多核基因来研究被子植物系统关系显得很重要。

虽然核基因在被子植物系统发育研究中的重要性不容忽视,但是编码蛋白质的核基因或核基因组的使用远远滞后于叶绿体基因的运用。一方面,由于核基因组拷贝低,导致使用传统的PCR扩增和测序得到序列信息相对困难,有时不得不依赖于克隆测序,工作量增大,限制了核基因在被子植物系统发育研究中的应用。另一方面,由于被子植物在进化过程中经历多次全基因组加倍(whole genome duplication)(Jiao *et al.*, 2011),近期的多倍化事件更加频繁,据保守估计被子植物中约有50%的物种是

由于多倍化形成的(Soltis *et al.*, 2009);其他机制导致的基因加倍事件也频繁发生(Ng & Yanofsky, 2001; Lin *et al.*, 2006; Xu *et al.*, 2009),导致不同类群中存在丢失不同拷贝的现象(Xu *et al.*, 2009)。这些因素使得同源核基因是直系同源还是旁系同源的关系难以确定。尽管如此,研究者们通过利用大量的测序资源与基因序列数据,开始在不同亲缘关系的类群间寻找可能是直系同源关系的单拷贝基因,并取得了一定进展。例如, Fulton等(2002)通过比较番茄(*Solanum lycopersicum*)的EST序列和拟南芥(*Arabidopsis thaliana*)的基因组序列,鉴定了两者共有的1,025个保守的低拷贝候选直系同源基因;Wu等(2006)比较番茄、土豆(*Solanum tuberosum*)、辣椒(*Capsicum annuum*)和咖啡(*Coffea arabica*)的EST序列与拟南芥的基因组序列,认为真菊类植物(euasterids)与拟南芥共有2,869个单拷贝候选直系同源基因;Duarte等(2010)鉴定了拟南芥、杨树(*Populus trichocarpa*)、葡萄(*Vitis vinifera*)和水稻(*Oryza sativa*)共有的959个单拷贝候选直系同源基因,并发现其中大部分基因在小立碗藓(*Physcomitrella patens*)和卷柏(*Selaginella moellendorffii*)中也能找到;Zhang等(2012)通过比较拟南芥、杨树、桃(*Prunus persica*)、葡萄、猴面花(*Mimulus guttatus*)、水稻、高粱(*Sorghum bicolor*)和小立碗藓的序列鉴定了1,083个高度保守的低拷贝核基因。这些工作为低拷贝核基因的利用奠定了基础。

然而,目前仍只有少数保守的核基因被用于研究被子植物各大类群间的亲缘关系。例如Mathews和Donoghue(1999)使用PHYA(Phytochrome A)和PHYC(Phytochrome C)推断Amborella是被子植物的最基部类群,是其他所有被子植物的姊妹分支(Mathews & Donoghue, 1999)。最近,Zhang等(2012)从分布在各大类群的73种被子植物中获得了5个十分保守的与染色体或核酸代谢相关的核基因序列,即SMC1(structure maintenance of chromosomes family protein), SMC2(structure maintenance of chromosomes family protein), MCM5(minichromosome maintenance family protein), MSH1(DNA mismatch repair MUTS related protein)和MLH1(DNA mismatch repair protein),发现它们的序列分化速度比叶绿体基因快,可以提供更多的系统发育信息;结合已完成全基因组测序的被子植物共91种(代表46

个目), 他们利用这些物种中5个核基因的信息重建了被子植物系统发育关系。他们的研究结果高度支持无油樟目(Amborellales)、睡莲目和木兰藤目(Austrobaileyales)为最基部的被子植物类群, 并且在多个层面与叶绿体基因组的研究结果总体保持一致, 即被子植物几大分支分别为单系类群, 如单子叶、真双子叶、木兰类、真双子叶中的蔷薇类和菊类, 以及各个目的划分等(图1; 细节请见第2节), 同时在许多之前有争议的分类位置上也得到了有高支持率的新结果。

在低等分类阶元系统发育重建过程中, 越来越多的研究者开始使用核基因或其内含子区域(Strand *et al.*, 1997; Sang, 2002; Mort & Crawford, 2004)。Ness等(2011)利用5个核基因重建了水生植物雨久花科(Pontederiaceae)的系统发育关系, 解决了叶绿体基因研究结果中凤眼莲属(*Eichhornia*)是并系(paraphyly)的问题, 并推断出凤眼莲(*E. meyeri*)位于该科的基部。Kim等(2008)利用*LEAFY*的第二个内含子阐明了蓼科(Polygonaceae)春蓼属(*Persicaria*)内复杂的杂交和多倍化事件。Álvarez等(2008)在菊科千里光族(Senecioneae)中筛选出了9个候选的直系同源基因, 并通过实验获得基因序列, 验证了*CesA*(cellulose synthase)、*CHS*(chalcone synthase)、*DHS*(DAHP synthase)和*QG8140*这4个核基因的可用性。Small和Wendel(2000)、Álvarez等(2005)利用核基因*Adh*(alcohol dehydrogenase)、*A1341*和*CesA1b*研究了棉属(*Gossypium*)内的物种关系。其他众多核基因也被用于研究不同类群的亲缘关系, 包括B类MADS-box 基因家族中的*AP3*(*APETALA3*)和*PI*(*PISTILLATA*)(Bailey & Doyle, 1999; Aoki *et al.*, 2004), *Adc*(arginine decarboxylase)(Galloway *et al.*, 1998), *RPB2*(RNA polymerase II)(Denton *et al.*, 1998), *Waxy*(Mason-Gamer *et al.*, 1998; Fortune *et al.*, 2007), *GAP3DH*(glyceraldehyde-3-phosphate dehydrogenase)(Gouy & Li, 1989), *ACCase*(acetyl-CoA carboxylase)(Huang *et al.*, 2002), *PGK*(plastid 3-phosphoglycerate kinase)(Huang *et al.*, 2002), *petD*(Löhne & Borsch, 2005), *GBSSI*(granule-bound starch synthase)(Mason-Gamer *et al.*, 1998; Evans *et al.*, 2000), *Gpat*(glycerol-3-phosphate acyltransferase)(Tank & Sang, 2001), *nepGS*(Emshwiller & Doyle, 1999), *GIGANTEA*(Fowler *et al.*, 1999),

*GPAI*(G protein  $\alpha$  subunit)(Ma *et al.*, 1990), *GAI*(GA Insensitive)(Wen *et al.*, 2007), *AGBI*(G protein  $\beta$  subunit)(Weiss *et al.*, 1994)和*PPR*(pentatricopeptide repeat)(Yuan *et al.*, 2009, 2010)等。另外, 为了推动核基因在被子植物系统发育研究中的应用, 研究者们在豆科、菊科和唇形科等科中开展了寻找适用的核基因, 并设计可用于研究相应科的系统发育通用引物的工作(Choi *et al.*, 2006; Chapman *et al.*, 2007; Curto *et al.*, 2012; Krak *et al.*, 2012)。

虽然核基因在被子植物系统发育研究中越来越被重视, 但是被子植物的遗传背景复杂, 在不同类群间, 不同目、科和属间核基因的拷贝数和进化速率都具有丰富的多样性, 使得目前大部分核基因的应用仍然面临着适用范围窄、引物通用性不高的问题。例如Steele等(2008)在蔷薇类亲缘关系较远的葫芦科(Cucurbitaceae)和牻牛儿苗科(Geraniaceae)检测了141个低拷贝核基因片段的可用性, 发现其中仅3个适用于葫芦科的系统发育研究, 2个适用于牻牛儿苗科, 没有一个能同时适用于这两个科。另外, 目前用于不同类群亲缘关系研究的核基因的选择往往有一定的主观性, 适用于不同分类阶元的标记基因筛选标准也迟迟没有提出, 究竟怎样的核基因适于作为通用标记基因仍然是一个亟待解决的难题。被子植物基因组和转录组(transcriptome)测序计划的开展, 将极大地推动更多的核基因组和转录组数据应用于寻找不同分类阶元植物共有的直系同源基因。根据我们的经验, 进化速率较快的低拷贝核基因可以提供丰富的信息位点, 从而可能作为低等分类阶元(科、属和种)的分子标记; 进化速率慢的低拷贝核基因比较保守, 可用于研究高等分类阶元或分化很早的类群间的亲缘关系。筛选出此类分子标记, 将大大提高分子系统学研究的效率, 改善现在通用核基因短缺的局面。

## 2 被子植物主要类群内已经确定的系统发育关系

从Hamby和Zimmer(1992)首次用18S和26S rDNA序列片段构建被子植物系统发育树, 到1998年被子植物系统发育组基于*rbcL*, *atpB*和18S rDNA的序列首次提出APG分类标准并不断修改更新至APG III(Bremer *et al.*, 2009), 以及最近运用编码蛋白的核基因序列研究各分类阶元被子植物的系统发育

关系, 人们对被子植物各类群间的进化关系有了全新的认识(Hamby & Zimmer, 1992; Jansen *et al.*,

2007; Moore *et al.*, 2007; Bremer *et al.*, 2009; Zhang *et al.*, 2012)(图1)。

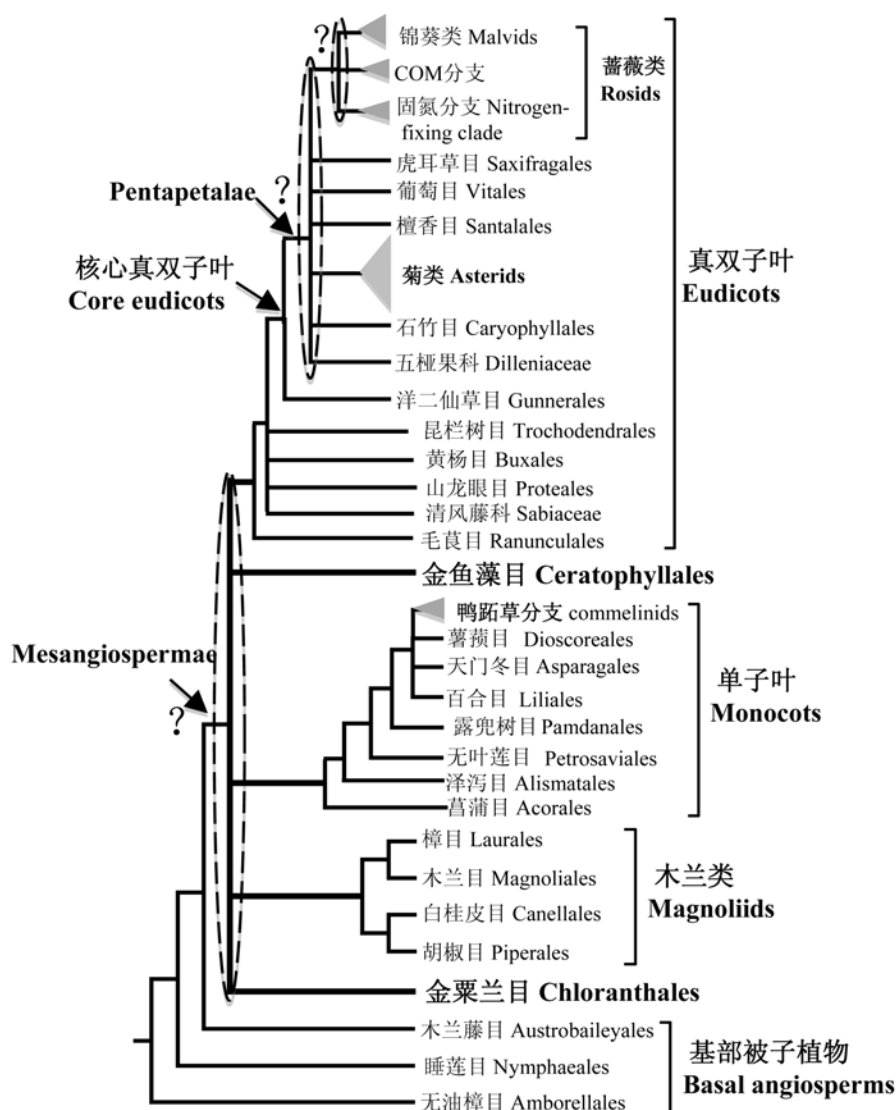


图1 被子植物各大类群关系图。除了位于最基部的3个目之外, 其余所有被子植物形成单系群。这个单系群由5个类群组成: 金粟兰目、木兰类、单子叶植物、金鱼藻目和真双子叶植物, 但是这5个分支的关系没有被确定, 因此五者的关系在树上以梳子结构来表示。单子叶植物中, 菖蒲目被认为是最基部的类群; 鸭跖草分支则是其核心类群, 包含4个目, 在树上以三角形表示。木兰类4个目的关系已经确定。毛茛目被认为位于真双子叶植物的最基部; 而在核心真双子叶植物中, 洋二仙草目被认为位于最基部, 其他的类群形成一个被称为Pentapetalae的单系群, 蔷薇类和菊类是其核心类群, 但是Pentapetalae内部各个类群的关系并没有被最终确定。蔷薇类中, COM分支与固氮分支还是锦葵类形成姊妹群仍然有争议。

Fig. 1 The relationships of major lineages of angiosperms. Except for three basal orders, all angiosperms formed a monophyletic group composed of five major lineages: Chloranthaceae, magnoliids, monocots, Ceratophyllaceae and eudicots. However, their relationships are not clear. Among monocots, Acorales is the basal order and commelinids (shown as triangle in this figure) with four orders are the core groups. The relationships among four magnoliid orders have been clarified. In eudicots, Ranunculales is sister to all other eudicots and Gunnerales is sister to all other core eudicots, which formed a monophyletic group known as Pentapetalae. Rosids and asterids are core groups of Pentapetalae, but there are many uncertain relationships in Pentapetalae. In rosids, it is still uncertain whether the sister group of the COM clade is the nitrogen-fixing clade or malvids.

## 2.1 被子植物的基部类群

目前被子植物系统发育框架已经基本形成,已知的现存物种隶属于哪个类群已经基本清楚,另外无油樟目、睡莲目和木兰藤目依次被认为是现存被子植物的最基部类群,也被称为“ANITA”(即 Amborellaceae, Nymphaeaceae, Illiciaceae, Trimeniaceae和Austrobaileyaceae的缩写)(Qiu *et al.*, 1999; Bremer *et al.*, 2009)。大多数研究表明,无油樟目是现存被子植物中最先分化出来的类群,此目仅含1个物种(*Amborella trichopoda*),为单科单属种,现存的自然群落只在新喀里多尼亚发现(New Caledonia)。但有少数研究结果也把无油樟目和睡莲目形成姊妹群放在被子植物最基部(Barkman *et al.*, 2000; Soltis & Soltis, 2004; Qiu *et al.*, 2005)。从形态学上来看,上述基部被子植物的种子都具2枚子叶。从系统发育树上,单子叶植物作为单系群嵌在传统意义上的双子叶植物内部,因此,传统意义上的双子叶植物就不再是单系群。这些结果也说明被子植物祖先的子叶数目可能是2枚,而后在单子叶植物起源前后丢失成为1枚,其他被子植物类群则保留了祖先中子叶的数目。

除了ANITA基部类群外,约占被子植物99.95%的其他类群聚为单系,此类群包括数量不等的5个分支,即真双子叶植物、单子叶植物(monocots)、木兰类(magnoliids)、金鱼藻目(Ceratophyllales)和金粟兰目(Chloranthales)(Cantino *et al.*, 2007)。真双子叶植物为五个分支中物种多样性最丰富的类群,包含了约40个目300余科,约占整个被子植物的75%,该类群的共同形态特征为花粉粒具有三孔沟(tricolpate pollen),因此也被称为“三孔花粉组”(Doyle & Hotton, 1991)。真双子叶植物起源古老,在中国发现的早白垩纪化石中就存在该类群植物(Drinnan *et al.*, 1991; Sun *et al.*, 2011),该类群的主要物种在约八千多万年前晚白垩纪桑托期(late Santonian)就已经形成(Herendeen *et al.*, 1999)。

## 2.2 真双子叶植物的系统发育关系

真双子叶植物中,毛茛目(Ranunculales)位于最基部,是其他所有真双子叶植物的姊妹群;清风藤科(Sabiaceae)、山龙眼目(Proteales)、黄杨目(Buxales)和昆栏树目(Trochodendrales)是真双子叶植物的其他基部类群;其余所有真双子叶物种形成单系,被命名为核心真双子叶(core eudicots)植物(Magallón

*et al.*, 1999)。在核心真双子叶植物中,洋二仙草目(Gunnerales)是其他类群的姊妹群;其他类群的花多是五基数,因此也被称为Pentapetalae,主要由蔷薇类(rosids)和菊类植物(asterids)两大类群组成,另外还包含系统位置尚未完全确定的几个类群,即虎耳草目(Saxifragales)、葡萄目(Vitales)、檀香目(Santalales)、五桠果科(Dilleniaceae)和石竹目(Caryophyllales)(Judd & Olmstead, 2004; Worberg *et al.*, 2007; Moore *et al.*, 2010)。

在蔷薇类植物中,主要类群为固氮分支(nitrogen-fixing clade),包括葫芦目(Cucurbitales)、壳斗目(Fagales)、蔷薇目(Rosales)和豆目(Fabales),是支持率很高的单系群;另外两个高支持率的分支是COM分支和锦葵类(malvids)分支,前者包括卫矛目(Celastrales)、酢浆草目(Oxalidales)和金虎尾目(Malpighiales),后者包括锦葵目(Malvales)、包含模式植物拟南芥的十字花目(Brassicales)、燧体木目(Crossosomatales)、无患子目(Sapindales)、十齿花目(Huerteales)和美洲苦木目(Picramniales)(Bremer *et al.*, 2009; Finet *et al.*, 2010; Qiu *et al.*, 2010; Zhang *et al.*, 2012)(图2a)。在菊类植物中,大部分物种组成一个被称为真菊的单系,而山茱萸目(Cornales)和杜鹃花目(Ericales)则是真菊的姊妹群(Bremer *et al.*, 2009; Zhang *et al.*, 2012)(图2b)。APG III又将真菊植物分为真菊I类(也被称为唇形类lamiales)和真菊II类(也被称为桔梗类campanulids);其中真菊I类的核心类群是唇形目(Lamiales)、茄目(Solanales)、龙胆目(Gentianales)、紫草科(Boraginaceae),而真菊II类的核心类群是盔瓣花目(Paracryphiales)、川续断目(Dipsacales)、伞形目(Apiiales)、鳞叶树目(Bruniales)、菊目(Asterales)和南鼠刺目(Escalloniales)。

## 2.3 单子叶植物的系统发育关系

单子叶植物是五大分支中的第二大类群,约占整个被子植物的22%,为人类提供了主要的粮食作物(谷物与其他淀粉类作物等)(Bremer, 2000; Stevens, 2001)。该类植物绝大多数为草本,种子仅有1枚子叶,平行叶脉,花基数为3。单子叶植物包含11个目和1个科(Bremer *et al.*, 2009),其中兰科植物约占单子叶物种的34%,禾本科植物约占17%。几乎所有的分子证据都表明菖蒲目(Acorales)是单子叶植物的最基部分支,泽泻目(Alismatales)紧随其后;另外,单子叶植物还包含天门冬目(Aspara-

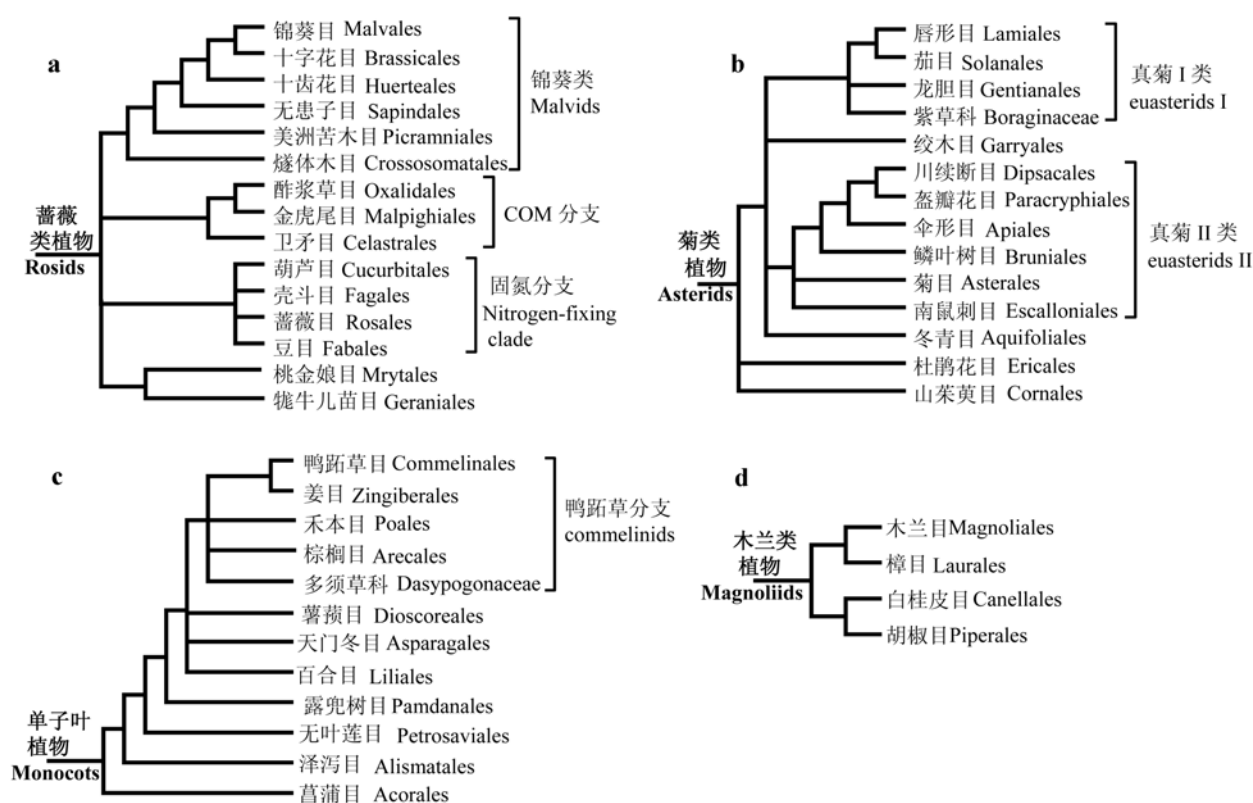


图2 蔷薇类植物、菊类植物、单子叶植物和木兰类植物内部关系图。(a)蔷薇类主要包含3个高支持率的分支,即锦葵类植物、COM分支和固氮分支,但是COM分支的姊妹群是固氮分支还是锦葵类植物仍然有争议,因此在图中以梳齿结构显示。(b)菊类植物主要分为真菊I类和II类,杜鹃花目和山茱萸目位于其基部;真菊I类内部的关系尚不清楚;绞木目和冬青目曾被认为是分别位于真菊I类和II类的基部,但最近来自核基因的研究结果表明二者以姊妹群形式位于真菊I类的基部。(c)单子叶植物的核心类群是鸭跖草分支,但该分支内部几个目的关系并不是很清楚;菖蒲目是单子叶植物的最基部类群。(d)木兰类植物中,木兰目与樟目、白桂皮目与胡椒目互为姊妹群。

Fig. 2 The relationships within rosids, asterids, monocots and magnoliids, respectively. (a) Rosids contain three clades: malvids, COM and the nitrogen-fixing clade, but whether the sister group of the COM clade is the nitrogen-fixing clade or malvids remains unclear. (b) Asterids are composed of euasterid I, euasterid II and two basal orders: Ericales and Cornales. The relationships among euasterids I are not clear. Garryales and Aquifoliales were respectively regarded as the basal group of euasterids I and II; however, new evidence based on nuclear genes suggested that these two together formed a sister group of other euasterids I. (c) Commelinids is the core group of monocots, and Acorales is the basal order. The relationships among commelinids are not clear. (d) In magnoliids, Magnoliales and Laurales are sister groups, and Canellales and Piperales are sister groups.

gales)、薯蓣目(Dioscoreales)、百合目(Liliales)、露兜树目(Pandanales)、无叶莲目(Petrosaviales)和鸭跖草类植物(commelinids)。其中鸭跖草类植物是单子叶植物的核心类群,包括棕榈目(Arecales; 木本)、鸭跖草目(Commelinales)、禾本目(Poales)、姜目(Zingiberales)和多须草科(Dasypogonaceae) (Davis *et al.*, 2004; Graham *et al.*, 2006)(图2c)。单子叶植物是被子植物各大类群中系统发育研究得比较清楚的类群之一,除了兰科和禾本科外,几乎其他单子叶植物所有属都曾进行了系统发育研究,大部分研

究基于叶绿体 $rbcL$ 基因,这为研究单子叶植物各分支的起源、扩张、重要形态性状的演化与适应等奠定了基础(Soltis *et al.*, 2005)。

## 2.4 木兰类植物、金粟兰目和金鱼藻目植物的系统发育关系

木兰类植物是五大分支中的第三大类群,现存物种多为常绿木本,包含白桂皮目(Canellales)、胡椒目(Piperales)、樟目(Laurales)和木兰目(Magnoliales),其中白桂皮目与胡椒目,樟目与木兰目互为姊妹群(图2d)。该类植物包含很多香料作

物,如胡椒(*Piper nigrum*)、肉豆蔻(*Myristica fragrans*)和香樟(*Cinnamomum camphora*)等(Bremer *et al.*, 2009)。

金鱼藻目和金粟兰目这两个小类群十分特别,除金粟兰目雪香兰(*Hedyosum*)的雌花外,两类植物的其他种都没有花被。金粟兰目只有1科4属70余种,只生长在热带和亚热带,为芳香植物,花很简单,有些种的花甚至被认为是被子植物最简单的花(Stevens, 2001)。金鱼藻目只有1科1属约5种,是全球广布的水生草本植物,其形态特征比较特殊,例如植株无根,叶轮生且边缘有散生的刺状细齿,花极细小不显眼(Stevens, 2001; Piep *et al.*, 2007)。

另外,除了上述五大类群之外,被子植物较低等分类阶元的研究也获得了重要进展,例如目水平的真双子叶的金虎尾目(Malpighiales)(Xi *et al.*, 2012)、虎耳草目(Saxifragales)(Fishbein *et al.*, 2001)和石竹目(Caryophyllales)(Cuénoud *et al.*, 2002; Crawley & Hilu, 2012)等,科水平的单子叶植物的禾本科(Tang *et al.*, 2010; Zhang *et al.*, 2011)和雨久花科(Ness *et al.*, 2011)等;属水平的真双子叶植物菊科(Asteraceae)的大吴风草属(*Farfugium*)(Nomura *et al.*, 2010)和兔儿风属(*Ainsliaea*)(Mitsui *et al.*, 2008)等。

### 3 被子植物系统发育研究中存在的问题

被子植物系统发育研究历经20多年的发展,虽然取得了巨大进展,但是仍然有很多系统发育关系存在争议,一些技术问题也亟待解决。

#### 3.1 被子植物主要类群中尚未确定的亲缘关系

被子植物在起源和漫长演化过程的不同时期中经历了若干次不同程度的扩张,有些分支在较短地质年代内爆发式形成大类群,导致了被子植物各个种群大小不一,而且产生了丰富的物种多样性(Davies *et al.*, 2004)。研究表明,除了5个分支形成时经历了快速物种扩张之外(Moore *et al.*, 2007),现存真双子叶的主要类群也源自至少1次物种扩张事件,蔷薇类(Wang *et al.*, 2009)、金虎尾目(Xi *et al.*, 2012)、虎耳草目(Fishbein *et al.*, 2001; Jian *et al.*, 2008)和石竹目(Cuénoud *et al.*, 2002)等在演化过程中也都经历了物种扩张。被子植物这些古老的物种扩张事件使各类群在形态和分子序列上的共有祖征较少,系统发育信号较少,从而导致物种亲缘关

系难以解决(Rokas & Carroll, 2006),在进化树上表现为“梳齿”(polytomies)结构。

首先,被子植物现存各大类群间在物种数量与形态特征上差异都很大,真双子叶植物、单子叶植物、木兰类、金鱼藻目和金粟兰目这5个类群间的关系一直存在争议(图1)。相比于其他3个类群,金鱼藻目和金粟兰目的物种数量很少,而且金鱼藻目为较特化的草本植物,往往在进化树上表现为长枝,这些因素都影响了5个类群间亲缘关系的解决。

关于这5个类群的亲缘关系,不同研究者根据不同基因或不同数目的物种得到的结果不尽相同,目前已发表的拓扑结构约有15种,但支持率都不高(Soltis *et al.*, 1999; Jansen *et al.*, 2007; Moore *et al.*, 2011; Zhang *et al.*, 2012)。其中,有研究分别将单子叶植物、单子叶植物与金鱼藻、金鱼藻和金粟兰、木兰类、金粟兰、木兰类与金粟兰作为这5个分支中的最基部类群。例如, Qiu等(1999)联合2个叶绿体基因、2个线粒体基因和18S rDNA的序列构建了包含105个种的系统树,认为真双子叶植物与木兰类植物是姊妹群,金粟兰目是两者的姊妹群,而单子叶植物与金鱼藻目形成的姊妹群位于最基部。有一些源于核基因、线粒体和叶绿体基因的证据也曾认为金鱼藻目与单子叶植物是姊妹群(Soltis *et al.*, 1997; Savolainen *et al.*, 2000),但金鱼藻目是比较特化的类群,而单子叶植物多为草本,进化速率快,两者在进化树上都表现为长枝,因此有研究者认为二者聚为姊妹群很可能是因为长枝吸引(long-branch attraction, LBA)(Soltis *et al.*, 2005)。

有些来自化石方面的资料却似乎仍表明金鱼藻目可能是比较原始的类群。例如在早期化石中发现的角状果(horned fruit)与金鱼藻的果实很像(Dilcher, 1989);另外被子植物最早的化石辽宁古果(*Archaeofructus liaoningensis*)的花很简单,这与现存的基部被子植物睡莲目的独蕊草科(Hydattellaceae)、金粟兰和金鱼藻的花很类似,因此有研究者推测可能金鱼藻目与金粟兰目的关系比较近,且位置都偏基部(Eklund *et al.*, 2004; Friis *et al.*, 2007; Endress & Doyle, 2009)。但是, Moore等(2007)基于叶绿体基因组的结果以及APG III,认为金鱼藻目是真双子叶植物的姊妹群,然后是单子叶植物,而木兰类与金粟兰目形成的姊妹群是五大分支最基部的类群。

除了化石证据之外(Friis *et al.*, 1986),现存金

粟兰目植物保留了很多祖先性状,例如瓶状心皮(ascidiate carpel)、梯纹导管孔(scalariform vessel perforation)、unilacunar nodal anatomy和chloranthoid leaf teeth,这些性状也曾让分类学家认为金粟兰目位于被子植物的基部(Swamy, 1953; Kong, 2001; Carpenter, 2006)。木兰类植物也保留了较多的被子植物祖先性状,如雄蕊多数、花粉多为单萌发孔、离生心皮等(Stevens, 2001),因此早期分类学家曾认为木兰科(Magnoliaceae)和番荔枝科(Annonaceae)等木本的木兰类植物很原始(Cronquist, 1968)。但是某些基部真双子叶植物中也具有很多被子植物的祖先性状,例如毛茛科植物也有精油细胞(ethereal oil cell)、无导管、镊合状的花药开裂(valvate anther dehiscence)和金粟兰叶齿形(chloranthoid leaf teeth)(Thorne, 1992; Drinnan *et al.*, 1994)。所以,单从形态特征来看,根据保留的祖先性状难以推测五大分支中的最基部类群。这5大分支的亲缘关系仍将是今后被子植物系统发育研究中需要重点关注的问题。

其次,除了被子植物5个主要分支的亲缘关系还不能确定外,各大类群内部主要分支的关系也没有完全厘清(图2)。真双子叶植物中,除了最基部的毛茛目外,其他基部类群中的清风藤科与山龙眼目、黄杨目与昆栏树目是否互为姊妹群,尚需进一步的研究(Moore *et al.*, 2010; Zhang *et al.*, 2011)。核心真双子叶植物中除了蔷薇类和菊类植物两大分支外,其他重要分支,即虎耳草目、檀香目、葡萄目、五桠果科和石竹目的系统发育位置也还没有最终确定(图1)。蔷薇类内部的很多关系也有争论(Finet *et al.*, 2010; Moore *et al.*, 2010)(图2a),例如COM分支曾被认为与固氮分支形成姊妹群,这两支共同被称为真蔷薇I类,但是最近利用5个核基因得到的具有高度支持的结果却表明,COM分支中的卫矛目和金虎尾目与锦葵类植物(原来认为属于真蔷薇II类)形成姊妹群,这与线粒体基因的研究结果一致(Zhu *et al.*, 2007; Finet *et al.*, 2010; Qiu *et al.*, 2010; Zhang *et al.*, 2012)。菊类植物中(图2b),叶绿体证据大都支持山茱萸目和杜鹃花目依次为其他菊类植物的姊妹群,但近来基于核基因的结果表明两者互为姊妹群,并且支持率高达100%,而两者合并的一支则与其他真菊类植物为姊妹群;绞木目(Garryales)和冬青目(Aquifoliales)曾被认为分别是真菊I和II类的最基部类群,但是近来核基因的结果

则认为二者互为姊妹群,并位于原来真菊I类的基部(Zhang *et al.*, 2012)。单子叶植物中,除了已知菖蒲目和泽泻目依次位于基部外,鸭跖草分支内部4个目1个科的关系尚不明晰;APG III认为天门冬目与鸭跖草分支关系最近,而后依次是百合目、薯蓣目+露兜树目以及无叶莲目,但是Zhang等(2012)利用5个核基因得到的结果却认为薯蓣目可能与鸭跖草分支靠近(图2c)。

第三,相比于高等分类阶元,被子植物低等分类阶元如目、科和属水平未解决的问题则更多。例如菊科(Asteraceae或Compositae)是真双子叶植物中最大的科,约1,600属,占开花植物属的10%(Funk *et al.*, 2005; Anderberg *et al.*, 2007)。按照头状花序中小花的构造以及植物有无乳汁等特征,菊科分为舌状花亚科(Liguliflorae)和管状花亚科(Tubuliflorae),但也有研究将菊科分为12个亚科,其中前4个较大的亚科为Asteroideae(有16,000多种,如青蒿、向日葵),Cichoroideae(3,200种,如莴苣),Carduoideae(2,500种,如大蓟)和Mutisioideae(750种)(Panero & Funk, 2008)。另外还有其他一些属的位置无法确定(Anderberg *et al.*, 2007),各亚科之间以及亚科内也有很多关系尚未确定。十字花科(Brassicaceae)中有44个族(tribe)307属约3,600种,被分为3个主要的系(lineage),但是还有约38个属和71个种无法确定属于哪个系或哪个族;3个系间的亲缘关系及系内部各族间和各属间的问题也仍然很多(Bailey *et al.*, 2006; Couvreur *et al.*, 2010; Franzke *et al.*, 2011)。

### 3.2 基于核苷酸或氨基酸序列的研究方法存在的问题及可能的解决办法

系统发育研究中,可能导致结果不准确的原因主要分为如下三类(Jeffroy *et al.*, 2006; Philippe *et al.*, 2011): (1)使用的分子标记不是真正的直系同源基因,即hidden paralogs。被子植物中全基因组加倍和其他基因重复事件多次发生,基因重复后不同拷贝的丢失事件使得直系同源基因的鉴定变得复杂(Jiao *et al.*, 2011)。(2)用于系统发育分析的基因长度较短或者信息含量较低时,“噪音”信息例如回复突变(reverse mutation)和平行突变(parallel mutation, 也被称为homoplasy)等,可能会掩盖真实的系统发育信息,从而可能引入随机误差。(3)不同物种中核苷酸成分或碱基替代模式的偏差(compositional signal)、类群之间进化速率的差异(rate signal)以及

特定碱基位点替换速率的差异(heterotachous signal)等因素造成的系统误差(systematic error)(Lopez *et al.*, 2002; Collins *et al.*, 2005)。由于世代长短、居群大小和核苷酸替换速率差异等因素, 不同被子植物类群的进化速率是不同的, 当类群间进化速率差异很大时, 进化快的类群(在进化树上表现为长枝)可能因为多重突变在某些碱基位点随机获得了相同的碱基, 从而使亲缘关系较远的物种在序列的部分位置上反而相似, 当这些非同源相似过多而足以掩盖真实的同源信息时, 就会错误地将进化快的类群聚在一起(即长枝吸引)(Goremykin *et al.*, 2003)。另外, 虽然目前常用的序列替代模型已经考虑了基因内部不同碱基位点因选择压力不同而有相异的进化速率, 但是几乎所有的模型都没有考虑选择压力会随时间发生变化, 即假设特定碱基位点的替代速率随着时间推移在类群之间保持恒定, 这与实际的分子进化过程并不一定相符(Yang, 1994; Gu *et al.*, 1995; Lopez *et al.*, 2002)。

目前被子植物系统发育研究主要分为两大类: 第一类是物种数目很多但基因数目较少(Burleigh *et al.*, 2009; Qiu *et al.*, 2010; Soltis *et al.*, 2011)。当选取的基因序列较短或包含的系统发育信息少时, 并不能解决所有关注的问题, 或者很可能产生随机误差。降低随机误差的方法是选择系统发育信息充足的直系同源基因, 加大基因数据量(Wortley *et al.*, 2005; Jian *et al.*, 2008), 但是当物种数目太多时, 每增加一个基因片段信息都会增加工作量和计算时间。第二类研究是物种数量较少但基因数目很多, 即系统发育基因组学(Delsuc *et al.*, 2005)。被子植物的叶绿体系统发育基因组学在近几年取得了很大的进展, 例如Jansen等(2007)用81个叶绿体基因构建包含64个物种的系统树, Moore等(2007)用61个叶绿体基因(约42,000 bp)构建包含45个物种的系统树。

系统发育分析中随着数据量的增加虽然能尽可能降低随机误差带来的影响, 但是却可能引入系统误差; 这些误差在某些情况下还可能被放大, 从而导致产生统计学上支持率很高但却不能反映真实物种关系的进化树(Phillips *et al.*, 2004; Jeffroy *et al.*, 2006)。因此, 相对于使用较少基因序列构建系统发育树, 用系统发育基因组学来研究物种间的亲缘关系时更需要谨慎(Soltis *et al.*, 2004)。被子植物

系统发育基因组学研究中比较典型的反例即由系统误差带来的长枝吸引假象(Philippe *et al.*, 2011), 其典型原因就是取样不当或取样量不够。被子植物基部类群的确定就出现过类似的问题。虽然很多研究表明*Amborella*或*Amborella*和*Nymphaeales*形成的姊妹群是被子植物的最基部类群, 但是Goremykin等(2003)测定*Amborella trichopoda*和*Nymphaea alba*的叶绿体全基因组序列后, 结合公共数据库中12个叶绿体基因组序列, 提出单子叶植物(禾本科为代表物种)为被子植物的最基部类群。然而Soltis和Stefanovic等发现在增加关键类群, 尤其是基部类群如*Nymphaea*、*Austrobaileya*、木兰类或基部单子叶植物后, *Amborella*或*Amborella* + *Nymphaeales*仍然是被子植物的最基部类群(Soltis & Soltis, 2004; Stefanović *et al.*, 2004), 这一结果在增加取样量后的叶绿体系统发育基因组学研究结果中也得到了验证(Jansen *et al.*, 2007; Moore *et al.*, 2007)。Goremykin等(2003, 2013)的结果很大程度上是由于取样不当或不够导致的, 禾本科的进化速率很快, 因此当仅以长枝的禾本科植物作为单子叶植物的代表, 以及选用买麻藤属(*Gnetum*)植物作为外类群时, 就会造成单子叶被吸引到基部的假象。除了选取合适的进化模型和分析方法之外, 密集采样以减少长枝, 目前被认为是降低长枝吸引影响的有效方法(Graybeal, 1998; Hillis *et al.*, 2003)。但是目前大部分实验室还是较难对物种密集测序, 而且当数据量增多时, 系统发育“噪音”也会增多, 从而可能放大系统误差。无论使用叶绿体基因组还是核基因组, 系统误差的影响都不容忽视。

综上所述, 通过仔细筛选直系同源基因(Zhang *et al.*, 2012), 选用适当的分子数据, 例如研究起源较早的类群之间的亲缘关系时使用氨基酸序列以减少密码子简并性带来的“噪音”, 或者使用系统发育信号强的基因降低干扰(Goremykin *et al.*, 2009; Townsend *et al.*, 2012; Salichos & Rokas, 2013), 使用核苷酸数据研究亲缘关系较远的类群时去除变化速率快的第三位密码子或位点(Jeffroy *et al.*, 2006; Goremykin *et al.*, 2010), 适当增加物种或选取进化较慢的物种(Graybeal, 1998), 开发新的符合基因序列进化过程的替换模型, 寻找其他序列特征例如插入缺失等策略来尽量降低各种误差的影响显得非常重要。

#### 4 基于核苷酸或氨基酸序列研究被子植物系统发育关系的未来趋势

随着被子植物系统发育学的发展,大量的研究数据涌现,TreeBASE(<http://www.treebase.org>)和Dryad (<http://datadryad.org>)等公共数据库即专门收录此类数据,以期促进植物系统学家间的数据交流。目前绝大多数系统发育研究都只包含了较少的物种,故许多研究者开始致力于构建包含大量乃至所有被子植物的系统发育树(“Big Tree”)。一种策略是涵盖GenBank中所有被子植物的种,另一种策略是构建属水平的被子植物系统树,例如目前GenBank中包含的被子植物约15,000属的系统树(Soltis & Soltis, 2013)。但是迄今为止,大多数的建树法仅能处理低于1,000条的序列,因此开发新的能处理超量数据的算法与分析软件显得至关重要。iPlant Collaborative(<http://www.iplantcollaborative.org>)即致力于开发构建大规模进化树与后续分析的工具(Soltis & Soltis, 2013)。最近Smith等(2011)成功利用RAxML对Genbank中55,473种的种子植物构建进化树,研究结果发现此超大树(mega-phylogeny)的整体拓扑结构与Soltis等(2011)包含640种被子植物的树的拓扑结构基本一致,表明依赖进化模型(model-based)的建树方法适用于构建被子植物的“Big Tree”。高效快速的超大型数据量分析软件的研发将为构建被子植物大系统树带来希望。届时,能定期依据GenBank新加入的物种更新被子植物大系统树软件的产生也将值得期待。

测序技术的发展使核基因组和转录组测序变得高效与快速,测序成本也大大降低,这为获得多个物种核基因的数据提供了坚实的技术基础,将使得联合大量基因乃至全基因组数据研究被子植物系统发育关系成为重要趋势,也将极大地促进核基因系统发育基因组学的研究。不同于利用PCR方法扩增少量基因,利用基因组或转录组测序能快速地获得大量的基因序列,可以寻找系统发育信息量和进化速率等适用于研究不同分类阶元亲缘关系的直系同源基因,应用于系统发育分析,常用的寻找直系同源基因的分析软件有InParanoid(O'Brien *et al.*, 2005), OrthoMCL(Li *et al.*, 2003)和HaMStR(Ebersberger *et al.*, 2009)等。动物中应用系统发育基因组学解决系统发育问题已经有了很多成功的先

例。例如Smith等(2012)获取了15种软体动物的转录组序列,联合1,185个基因重建了软体动物系统发育树;Kocot等(2011)测定了20种软体动物的转录组序列并联合基因组序列,利用HaMStR鉴定出308个共有的直系同源基因后,得到了高支持率的软体动物主要分支间的亲缘关系树;Johnson等(2013)利用基因组数据重建了膜翅目中社会性昆虫蚂蚁、蜜蜂和黄蜂的系统发育树,发现蚂蚁与蜜蜂总科(包含spheciform wasps和bees)是姊妹群。系统发育基因组学方法也逐步被应用于研究被子植物各类群的亲缘关系,如Wen等(2013)测定了葡萄科15个物种的转录组序列,联合229个单拷贝的直系同源基因序列,确定了葡萄科5个分支间的亲缘关系。

目前,系统发育基因组学的策略主要是超级矩阵(supermatrix)法和超级树(supertree)法。超级矩阵法是将所有比对后的单基因序列拼接(concatenate)成一个完整大矩阵后建树,在叶绿体基因中已经被成功应用(Jansen *et al.*, 2007; Moore *et al.*, 2007, 2010),但在核基因中的应用尚不多。华大基因研究院(BGI)正在开展测定1,000种植物转录组的计划(1KP Project)(<http://www.onekp.com/>),被子植物系统发育研究中一直难以解决的重要问题都有望利用核基因系统发育基因组学得到解决,例如被子植物5个主要分支之间的关系,真双子叶植物内部位置不确定类群的关系等。如前所述,超级矩阵法无法避免系统误差,很可能会得到统计上支持率很高的拓扑结构,实际却不能反映物种之间真实的亲缘关系,所以也需要谨慎选择基因。

超级树法即按照溯祖理论(coalescent model)(Kingman, 1982, 2000)基于大量单基因树推断物种树,常用的分析软件有BEST(Bayesian Estimation of Species Trees)(Liu, 2008)、GLASS(Global LateSt Split)(Liu *et al.*, 2010)、STAR(Species Tree Estimation using Average Ranks of Coalescences)(Liu *et al.*, 2009)、STEAC(Species Tree Estimation using Average Coalescence Times)(Liu *et al.*, 2009)和STEM(Species Tree Estimation Using Maximum Likelihood for Gene Trees under Coalescence)(Kubatko *et al.*, 2009)等。Burleigh等(2009)联合18,896个核基因树重建了136种植物的系统发育树,拓扑结构框架与前人基本一致,但是该结果并不支持基于叶绿体基因划分的真蔷薇I类和II类,这与其

他核基因的结果一致(Finet *et al.*, 2010; Zhang *et al.*, 2012); 并发现经历过基因重复和丢失事件的核基因可以通过超级树法构建系统发育树, 从而推动了核基因在系统发育研究中的应用(Burleigh *et al.*, 2009)。Shaw等(2013)开发了在线分析工具STRAW (Species TRee Analysis Web server), 大大方便了使用者。

另外, 二代测序技术和RAD(restriction-site associated DNA)技术相结合, 为群体基因组学提供了大量数据。最近有研究者将此技术用于低等分类阶元重建并取得了好的结果。例如Eaton和Ree(2013)在列当科(Orobanchaceae)植物中使用40,000个RAD位点鉴定了古老的杂交事件(Eaton & Ree, 2013); Wang等(2013)利用此方法鉴定了SNP分子标记, 并重建了温带竹子的系统发育关系。由此可见, 今后数年, 测序技术的发展将进一步加速植物基因数据的产生。结合新的分析方法与工具, 被子植物不同分类阶元系统发育关系, 乃至所有植物的进化关系, 将得到全面崭新的认识, 为植物科学的各个领域奠定坚实的基础。

**致谢:** 感谢孔宏智研究员、杨继教授、国春策、钱胜沾、王海峰、王雅琼和尤辰江对本文提出的宝贵意见。

## 参考文献

- Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution*, **29**, 417–434.
- Álvarez I, Cronn R, Wendel JF (2005) Phylogeny of the New World diploid cottons (*Gossypium* L., Malvaceae) based on sequences of three low-copy nuclear genes. *Plant Systematics and Evolution*, **252**, 199–214.
- Álvarez I, Costa A, Feliner GN (2008) Selecting single-copy nuclear genes for plant phylogenetics: a preliminary analysis for the Senecioneae (Asteraceae). *Journal of Molecular Evolution*, **66**, 276–291.
- Anderberg AA, Baldwin BG, Bayer RG, Breitwieser J, Jeffrey C, Dillon MO, Eldenas P, Funk V, Garcia-Jacas N, Hind DJN (2007) Compositae. In: *The Families and Genera of Vascular Plants* (ed. Kubitzki K), pp. 61–588, Springer.
- Aoki S, Uehara K, Imafuku M, Hasebe M, Ito M (2004) Phylogeny and divergence of basal angiosperms inferred from *APETALA-3* and *PISTILLATA*-like MADS-box genes. *Journal of Plant Research*, **117**, 229–244.
- APG (1998) An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden*, **85**, 531–553.
- APG (2003) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society*, **141**, 399–436.
- Bailey CD, Doyle JJ (1999) Potential phylogenetic utility of the low-copy nuclear gene *pistillata* in dicotyledonous plants: comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. *Molecular Phylogenetics and Evolution*, **13**, 20–30.
- Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick SI, Windham MD, Al-Shehbaz IA (2006) Toward a global phylogeny of the Brassicaceae. *Molecular Biology and Evolution*, **23**, 2142–2160.
- Baldwin BG, Sanderson MJ, Porter JM, Wojciechowski MF, Campbell CS, Donoghue MJ (1995) The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, **82**, 247–277.
- Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, Wolfe AD, dePamphilis CW (2000) Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. *Proceedings of the National Academy of Sciences, USA*, **97**, 13166–13171.
- Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, **424**, 197–201.
- Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, vos Rutger A, Gittleman JL, Purvis A (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.
- Bremer B, Bremer K, Chase M, Fay M, Reveal J, Soltis D, Soltis P, Stevens P (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*, **161**, 105–121.
- Bremer K (2000) Early Cretaceous lineages of monocot flowering plants. *Proceedings of the National Academy of Sciences, USA*, **97**, 4707–4711.
- Burleigh JG, Hilu KW, Soltis DE (2009) Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evolutionary Biology*, **9**, 61.
- Cantino PD, Doyle JA, Graham SW, Judd WS, Olmstead RG, Soltis DE, Soltis PS, Donoghue MJ (2007) Towards a phylogenetic nomenclature of Tracheophyta. *Taxon*, **56**, 1E–44E.
- Carpenter KJ (2006) Specialized structures in the leaf epidermis of basal angiosperms: morphology, distribution, and homology. *American Journal of Botany*, **93**, 665–681.
- Chapman MA, Chang JC, Weisman D, Kesseli RV, Burke JM (2007) Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae). *Theoretical and Applied Genetics*, **115**, 747–755.
- Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH,

- Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden*, **80**, 528–580.
- Choi HK, Luckow MA, Doyle J, Cook DR (2006) Development of nuclear gene-derived molecular markers linked to legume genetic maps. *Molecular Genetics and Genomics*, **276**, 56–70.
- Collins TM, Fedrigo O, Naylor GJP (2005) Choosing the best genes for the job: the case for stationary genes in genome-scale phylogenetics. *Systematic Biology*, **54**, 493–500.
- Corriveau JL, Coleman AW (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. *American Journal of Botany*, **75**, 1443–1458.
- Couvreux TL, Franzke A, Al-Shehbaz IA, Bakker FT, Koch MA, Mummenhoff K (2010) Molecular phylogenetics, temporal diversification, and principles of evolution in the mustard family (Brassicaceae). *Molecular Biology and Evolution*, **27**, 55–71.
- Crane PR, Lidgard S (1989) Angiosperm diversification and paleolatitudinal gradients in Cretaceous floristic diversity. *Science*, **246**, 675–678.
- Crane PR, Herendeen P, Friis EM (2004) Fossils and plant phylogeny. *American Journal of Botany*, **91**, 1683–1699.
- Crawley S, Hilu K (2012) Impact of missing data, gene choice, and taxon sampling on phylogenetic reconstruction: the Caryophyllales (angiosperms). *Plant Systematics and Evolution*, **298**, 297–312.
- Cronquist A (1968) *The Evolution and Classification of Flowering Plants*. Houghton Mifflin, Boston.
- Cuénoud P, Savolainen V, Chatrou LW, Powell M, Grayer RJ, Chase MW (2002) Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*, *atpB*, and *matK* DNA sequences. *American Journal of Botany*, **89**, 132–144.
- Curto MA, Puppo P, Ferreira D, Nogueira M, Meimberg H (2012) Development of phylogenetic markers from single-copy nuclear genes for multi locus, species level analyses in the mint family (Lamiaceae). *Molecular Phylogenetics and Evolution*, **63**, 758–767.
- Darwin F, Seward AC (1903) *More Letters of Charles Darwin*. Cambridge University Press, London, UK.
- Davies TJ, Barraclough TG, Chase MW, Soltis PS, Soltis DE, Savolainen V (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences, USA*, **101**, 1904–1909.
- Davis CC, Wurdack KJ (2004) Host-to-parasite gene transfer in flowering plants: phylogenetic evidence from Malpighiales. *Science*, **305**, 676–678.
- Davis JI, Stevenson DW, Petersen G, Seberg O, Campbell LM, Freudenstein JV, Goldman DH, Hardy CR, Michelangeli FA, Simmons MP, Mark P, Specht CD, Vergara-Silva F, Gandolfo M (2004) A phylogeny of the monocots, as inferred from *rbcL* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Systematic Botany*, **29**, 467–510.
- Day A, Ellis T (1984) Chloroplast DNA deletions associated with wheat plants regenerated from pollen: possible basis for maternal inheritance of chloroplasts. *Cell*, **39**, 359–368.
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nature Reviews Genetics*, **6**, 361–375.
- Denton AL, McConaughy BL, Hall BD (1998) Usefulness of RNA polymerase II coding sequences for estimation of green plant phylogeny. *Molecular Biology and Evolution*, **15**, 1082–1085.
- Dilcher DL (1989) The occurrence of fruits with affinities to Ceratophyllaceae in lower and mid-Cretaceous sediments. *American Journal of Botany*, **76**, 162.
- Doyle JA, Hotton CL (1991) Diversification of early angiosperm pollen in a cladistic context. In: *Pollen and Spores: Pattern of Diversification* (eds Blackmore S, Barnes SH), pp. 169–195. Clarendon Press, Oxford, England.
- Drinnan AN, Crane PR, Friis EM, Pedersen KR (1991) Angiosperm flowers and tricolpate pollen of buxaceous affinity from the Potomac Group (mid-Cretaceous) of eastern North America. *American Journal of Botany*, **78**, 153–176.
- Duarte JM, Wall PK, Edger PP, Landherr LL, Ma H, Pires JC, Leebens-Mack J, dePamphilis CW (2010) Identification of shared single copy nuclear genes in *Arabidopsis*, *Populus*, *Vitis* and *Oryza* and their phylogenetic utility across various taxonomic levels. *BMC Evolutionary Biology*, **10**, 61.
- Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq Data: an example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, **62**, 689–706.
- Ebersberger I, Strauss S, von Haeseler A (2009) HaMStR: Profile hidden markov model based search for orthologs in ESTs. *BMC Evolutionary Biology*, **9**, 157.
- Eklund H, Doyle JA, Herendeen PS (2004) Morphological phylogenetic analysis of living and fossil Chloranthaceae. *International Journal of Plant Sciences*, **165**, 107–151.
- Emshwiller E, Doyle JJ (1999) Chloroplast-expressed glutamine synthetase (npsGS): potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). *Molecular Phylogenetics and Evolution*, **12**, 310–319.
- Endress PK, Doyle JA (2009) Reconstructing the ancestral angiosperm flower and its initial specializations. *American Journal of Botany*, **96**, 22–66.
- Evans RC, Alice LA, Campbell CS, Kellogg EA, Dickinson TA (2000) The granule-bound starch synthase (GBSSI) gene in the Rosaceae: multiple loci and phylogenetic utility. *Molecular Phylogenetics and Evolution*, **17**, 388–400.
- Farrell BD (1998) “Inordinate fondness” explained: why are there so many beetles? *Science*, **281**, 555–559.
- Fauré S, Noyer JL, Carreel F, Horry JP, Bakry F, Lanaud C (1994) Maternal inheritance of chloroplast genome and paternal inheritance of mitochondrial genome in bananas

- (*Musa acuminata*). *Current Genetics*, **25**, 265–269.
- Finet C, Timme RE, Delwiche CF, Marlétaz F (2010) Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology*, **20**, 2217–2222.
- Fishbein M, Hibsch-Jetter C, Soltis DE, Hufford L (2001) Phylogeny of Saxifragales (angiosperms, eudicots): analysis of a rapid, ancient radiation. *Systematic Biology*, **50**, 817–847.
- Fortune P, Schierenbeck KA, Ainouche A, Jacquemin J, Wendel JF, Ainouche M (2007) Evolutionary dynamics of *Waxy* and the origin of hexaploid *Spartina* species (Poaceae). *Molecular Phylogenetics and Evolution*, **43**, 1040–1055.
- Fowler S, Lee K, Onouchi H, Samach A, Richardson K, Morris B, Coupland G, Putterill J (1999) GIGANTEA: a circadian clock-controlled gene that regulates photoperiodic flowering in *Arabidopsis* and encodes a protein with several possible membrane-spanning domains. *The EMBO Journal*, **18**, 4679–4688.
- Franzke A, Lysak MA, Al-Shehbaz IA, Koch MA, Mummenhoff K (2011) Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends in Plant Science*, **16**, 108–116.
- Friedman WE (2009) The meaning of Darwin's "abominable mystery". *American Journal of Botany*, **96**, 5–21.
- Friis E, Crane P, Pedersen K (1986) Floral evidence for Cretaceous chloranthoid angiosperms. *Nature*, **320**, 163–164.
- Friis EM, Crane PR, Pedersen KR, Bengtson S, Donoghue PC, Grimm GW, Stampanoni M (2007) Phase-contrast X-ray microtomography links Cretaceous seeds with Gnetales and Bennettitales. *Nature*, **450**, 549–552.
- Fulton TM, van der Hoeven R, Eannetta NT, Tanksley SD (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *The Plant Cell*, **14**, 1457–1467.
- Funk VA, Randall BA, Sterling-Keeley ER, Chan R, Watson L, Gemeinholzer B, Schilling E, Panero JL, Baldwin BG, Nuriá GJ (2005) Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologiske Skrifter*, **55**, 343–374.
- Galloway GL, Malmberg RL, Price RA (1998) Phylogenetic utility of the nuclear gene arginine decarboxylase: an example from Brassicaceae. *Molecular Biology and Evolution*, **15**, 1312–1320.
- Goremykin VV, Hirsch-Ernst KI, Wölfl S, Hellwig FH (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Molecular Biology and Evolution*, **20**, 1499–1505.
- Goremykin VV, Viola R, Hellwig FH (2009) Removal of noisy characters from chloroplast genome-scale data suggests revision of phylogenetic placements of *Amborella* and *Ceratophyllum*. *Journal of Molecular Evolution*, **68**, 197–204.
- Goremykin VV, Nikiforova SV, Bininda-Emonds OR (2010) Automated removal of noisy data in phylogenomic analyses. *Journal of Molecular Evolution*, **71**, 319–331.
- Goremykin VV, Nikiforova SV, Biggs PJ, Zhong B, Delange P, Martin W, Woetzel S, Atherton RA, Mclenachan PA, Lockhart PJ (2013) The evolutionary root of flowering plants. *Systematic Biology*, **62**, 50–61.
- Gouy M, Li W (1989) Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. *Molecular Biology and Evolution*, **6**, 109–122.
- Graham SW, Zgurski JM, McPherson MA, Cherniawsky DM, Saarela JM, Horne EF, Smith SY, Wong WA, O'Brien HE, Biron VL (2006) Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *Aliso*, **22**, 3–20.
- Graybeal A (1998) Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology*, **47**, 9–17.
- Gu X, Fu YX, Li WH (1995) Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Molecular Biology and Evolution*, **12**, 546–557.
- Hamby RK, Zimmer EA (1988) Ribosomal RNA sequences for inferring phylogeny within the grass family (Poaceae). *Plant Systematics and Evolution*, **160**, 29–37.
- Herendeen PS, Magallon-Puebla S, Lupia R, Crane PR, Kobylinska J (1999) A preliminary conspectus of the Allon flora from the late Cretaceous (late Santonian) of central Georgia, USA. *Annals of the Missouri Botanical Garden*, **86**, 407–471.
- Hillis DM, Pollock DD, McGuire JA, Zwickl DJ (2003) Is sparse taxon sampling a problem for phylogenetic inference? *Systematic Biology*, **52**, 124.
- Hilu KW, Liang HP (1997) The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany*, **84**, 830–830.
- Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R (2003) Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany*, **90**, 1758–1776.
- Huang SX, Sirikhachornkit A, Faris JD, Su XJ, Gill BS, Haselkorn R, Gornicki P (2002) Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Molecular Biology*, **48**, 805–820.
- Jansen RK, Cai ZQ, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Müller KF, Guisinger-Bellian M, Haberle RC, Hansen AK (2007) Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences, USA*, **104**, 19369–19374.
- Jansen RK, Saski C, Lee SB, Hansen AK, Daniell H (2011) Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Molecular Biology and Evolution*, **28**, 835–847.

- Jeffroy O, Brinkmann H, Delsuc F, Philippe H (2006) Phylogenomics: the beginning of incongruence? *Trends in Genetics*, **22**, 225–231.
- Jian SG, Soltis PS, Gitzendanner MA, Moore MJ, Li RQ, Hendry TA, Qiu YL, Dhingra A, Bell CD, Soltis DE (2008) Resolving an ancient, rapid radiation in Saxifragales. *Systematic Biology*, **57**, 38–57.
- Jiao YN, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang HY, Soltis PS (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.
- Johnson BR, Borowiec ML, Chiu JC, Lee EK, Atallah J, Ward PS (2013) Phylogenomics resolves evolutionary relationships among ants, bees, and wasps. *Current Biology*, **23**, 2058–2062.
- Judd WS, Olmstead RG (2004) A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany*, **91**, 1627–1644.
- Kim K, Jansen RK (1995) *ndhF* sequence evolution and the major clades in the sunflower family. *Proceedings of the National Academy of Sciences, USA*, **92**, 10379–10383.
- Kim ST, Sultan SE, Donoghue MJ (2008) Allopolyploid speciation in *Persicaria* (Polygonaceae): insights from a low-copy nuclear region. *Proceedings of the National Academy of Sciences, USA*, **105**, 12370–12375.
- Kingman JF (1982) On the genealogy of large populations. *Journal of Applied Probability*, **19**, 27–43.
- Kingman JF (2000) Origins of the coalescent: 1974–1982. *Genetics*, **156**, 1461–1463.
- Kocot KM, Cannon JT, Todt C, Citarella MR, Kohn AB, Meyer A, Santos SR, Schander C, Moroz LL, Lieb B, Halanych KM (2011) Phylogenomics reveals deep molluscan relationships. *Nature*, **477**, 452–456.
- Kong HZ (2001) Comparative morphology of leaf epidermis in the Chloranthaceae. *Botanical Journal of the Linnean Society*, **136**, 279–294.
- Krak K, Álvarez I, Caklová P, Costa A, Chrtek J, Fehrer J (2012) Development of novel low-copy nuclear markers for *Hieraciinae* (Asteraceae) and their perspective for other tribes. *American Journal of Botany*, **99**, e74–e77.
- Kubatko LS, Carstens BC, Knowles LL (2009) STEM: species tree estimation using maximum likelihood for gene trees under coalescence. *Bioinformatics*, **25**, 971–973.
- Lee EK, Cibrian-Jaramillo A, Kolokotronis SO, Katari MS, Stamatakis A, Ott M, Chiu JC, Little DP, Stevenson DW, McCombie WR (2011) A functional phylogenomic view of the seed plants. *PLoS Genetics*, **7**, e1002411.
- Li DZ, Gao LM, Li H, Wang H, Ge XJ, Liu JQ, Chen ZD, Zhou SL, Chen SL, Yang JB (2011) Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences, USA*, **108**, 19641–19646.
- Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Research*, **13**, 2178–2189.
- Lin ZG, Kong HZ, Nei M, Ma H (2006) Origins and evolution of the *recA/RAD51* gene family: evidence for ancient gene duplication and endosymbiotic gene transfer. *Proceedings of the National Academy of Sciences, USA*, **103**, 10328–10333.
- Liu L (2008) BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*, **24**, 2542–2543.
- Liu L, Yu LL, Pearl DK, Edwards SV (2009) Estimating species phylogenies using coalescence times among sequences. *Systematic Biology*, **58**, 468–477.
- Liu L, Yu LL, Pearl DK (2010) Maximum tree: a consistent estimator of the species tree. *Journal of Mathematical Biology*, **60**, 95–106.
- Löhne C, Borsch T (2005) Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. *Molecular Biology and Evolution*, **22**, 317–332.
- Lonsdale D, Brears T, Hodge T, Melville SE, Rottmann W (1988) The plant mitochondrial genome: homologous recombination as a mechanism for generating heterogeneity. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, **319**, 149–163.
- Lopez P, Casane D, Philippe H (2002) Heterotachy, an important process of protein evolution. *Molecular Biology and Evolution*, **19**, 1–7.
- Magallón S, Castillo A (2009) Angiosperm diversification through time. *American Journal of Botany*, **96**, 349–365.
- Magallón S, Crane PR, Herendeen PS (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Annals of the Missouri Botanical Garden*, **96**, 297–372.
- Ma H, Yanofsky MF, Meyerowitz EM (1990) Molecular cloning and characterization of *GPA1*, a G protein alpha subunit gene from *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences, USA*, **87**, 3821–3825.
- Mason-Gamer RJ, Weil CF, Kellogg EA (1998) Granule-bound starch synthase: structure, function, and phylogenetic utility. *Molecular Biology and Evolution*, **15**, 1658–1673.
- Mathews S, Donoghue MJ (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science*, **286**, 947–950.
- Meng SW, Chen ZD, Li DZ, Liang HX (2002) Phylogeny of Saururaceae based on mitochondrial *matR* gene sequence data. *Journal of Plant Research*, **115**, 0071–0076.
- Mitsui Y, Chen ST, Zhou ZK, Peng CI, Deng YF, Setoguchi H (2008) Phylogeny and biogeography of the genus *Ainsliaea* (Asteraceae) in the Sino-Japanese region based on nuclear rDNA and plastid DNA sequence data. *Annals of Botany*, **101**, 111–124.
- Moore MJ, Bell CD, Soltis PS, Soltis DE (2007) Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. *Proceedings of the National Academy of Sciences, USA*, **104**, 19363–19368.
- Moore MJ, Soltis PS, Bell CD, Burleigh JG, Soltis DE (2010) Phylogenetic analysis of 83 plastid genes further resolves the early diversification of eudicots. *Proceedings of the National Academy of Sciences, USA*, **107**, 4623–4628.
- Moore MJ, Hassan N, Gitzendanner MA, Bruenn RA, Croley

- M, Vandeventer A, Horn JW, Dhingra A, Brockington SF, Latvis M (2011) Phylogenetic analysis of the plastid inverted repeat for 244 species: insights into deeper-level angiosperm relationships from a long, slowly evolving sequence region. *International Journal of Plant Sciences*, **172**, 541–558.
- Moreau CS, Bell CD, Vila R, Archibald SB, Pierce NE (2006) Phylogeny of the ants: diversification in the age of angiosperms. *Science*, **312**, 101–104.
- Mort ME, Crawford DJ (2004) The continuing search: low-copy nuclear sequences for lower-level plant molecular phylogenetic studies. *Taxon*, **53**, 257–261.
- Ness RW, Graham SW, Barrett SC (2011) Reconciling gene and genome duplication events: using multiple nuclear gene families to infer the phylogeny of the aquatic plant family Pontederiaceae. *Molecular Biology and Evolution*, **28**, 3009–3018.
- Ng M, Yanofsky MF (2001) Function and evolution of the plant *MADS-box* gene family. *Nature Reviews Genetics*, **2**, 186–195.
- Nomura N, Takaso T, Peng CI, Kono Y, Oginuma K, Mitsui Y, Setoguchi H (2010) Molecular phylogeny and habitat diversification of the genus *Farfugium* (Asteraceae) based on nuclear rDNA and plastid DNA. *Annals of Botany*, **106**, 467–482.
- O'Brien KP, Remm M, Sonnhammer EL (2005) Inparanoid: a comprehensive database of eukaryotic orthologs. *Nucleic Acids Research*, **33**, D476–D480.
- Olmstead RG, Palmer JD (1994) Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany*, **81**, 1205–1224.
- Palmer JD, Herbon LA (1988) Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, **28**, 87–97.
- Palmer JD (1990) Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature*, **348**, 337–339.
- Panero JL, Funk VA (2008) The value of sampling anomalous taxa in phylogenetic studies: major clades of the Asteraceae revealed. *Molecular Phylogenetics and Evolution*, **47**, 757–782.
- Philippe H, Brinkmann H, Lavrov DV, Littlewood DTJ, Manuel M, Wörheide G, Baurain D (2011) Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology*, **9**, e1000602.
- Phillips MJ, Delsuc F, Penny D (2004) Genome-scale phylogeny and the detection of systematic biases. *Molecular Biology and Evolution*, **21**, 1455–1458.
- Piep MB, Gunn BF, Roberts C (2007) *Flora of North America*. Oxford University Press, New York.
- Qiu YL, Lee JH, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen ZD, Savolainen V, Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature*, **402**, 404–407.
- Qiu YL, Dombrowska O, Lee JH, Li LB, Whitlock BA, Bernasconi-Quadroni F, Rest JS, Davis CC, Borsch T, Hilu KW (2005) Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *International Journal of Plant Sciences*, **166**, 815–842.
- Qiu YL, Li L, Wang B, Xue JY, Hendry TA, Li RQ, Brown JW, Liu Y, Hudson GT, Chen ZD (2010) Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution*, **48**, 391–425.
- Roelants K, Gower DJ, Wilkinson M, Loader SP, Biju S, Guillaume K, Moriau L, Bossuyt F (2007) Global patterns of diversification in the history of modern amphibians. *Proceedings of the National Academy of Sciences, USA*, **104**, 887–892.
- Rokas A, Williams BL, King N, Carroll SB (2003) Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature*, **425**, 798–804.
- Rokas A, Carroll SB (2006) Bushes in the tree of life. *PLoS Biology*, **4**, e352.
- Salichos L, Rokas A (2013) Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature*, **497**, 327–331.
- Sang T (2002) Utility of low-copy nuclear gene sequences in plant phylogenetics. *Critical Reviews in Biochemistry and Molecular Biology*, **37**, 121–147.
- Savolainen V, Chase MW, Hoot SB, Morton CM, Soltis DE, Bayer C, Fay MF, De Bruijn AY, Sullivan S, Qiu YL (2000) Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcl* gene sequences. *Systematic Biology*, **49**, 306–362.
- Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R (2004) Ferns diversified in the shadow of angiosperms. *Nature*, **428**, 553–557.
- Shaw, Ruan TZ, Glenn T, Liu L (2013) STRAW: Species TRee Analysis Web server. *Nucleic Acids Research*, **41**, 238–241.
- Small RL, Wendel JF (2000) Phylogeny, duplication, and intraspecific variation of *Adh* sequences in new world diploid cottons (*Gossypium* L., Malvaceae). *Molecular Phylogenetics and Evolution*, **16**, 73–84.
- Smith SA, Beaulieu JM, Donoghue MJ (2010) An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences, USA*, **107**, 5897–5902.
- Smith SA, Beaulieu JM, Stamatakis A, Donoghue MJ (2011) Understanding angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany*, **98**, 404–414.
- Smith SA, Wilson NG, Goetz FE, Feehery C, Andrade SC, Rouse GW, Giribet G, Dunn CW (2012) Corrigendum: resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, **493**, 708–708.
- Soltis DE, Albert VA, Savolainen V, Hilu K, Qiu YL, Chase MW, Farris JS, Stefanović S, Rice DW, Palmer JD (2004) Genome-scale data, angiosperm relationships, and ‘ending incongruence’: a cautionary tale in phylogenetics. *Trends in Plant Science*, **9**, 477–483.

- Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Wall PK, Soltis PS (2009) Polyploidy and angiosperm diversification. *American Journal of Botany*, **96**, 336–348.
- Soltis DE, Smith SA, Cellinese N, Wurdack KJ, Tank DC, Brockington SF, Refulio-Rodriguez NF, Walker JB, Moore MJ, Carlswald BS, Bell CD, Latvis M, Crawley S, Black C, Diouf D, Xi Z, Rushworth CA, Gitzendanner MA, Sytsma KJ, Qiu YL, Hilu KW, Davis CC, Sanderson MJ, Beaman RS, Olmstead RG, Judd WS, Donoghue MJ, Soltis PS (2011) Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany*, **98**, 704–730.
- Soltis DE, Soltis PS, Endress PK, Chase MW (2005) Phylogeny and evolution of angiosperms. Sunderland, Massachusetts.
- Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF (2000) Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society*, **133**, 381–461.
- Soltis DE, Soltis PS, Nickrent DL, Johnson LA, Hahn WJ, Hoot SB, Sweere JA, Kuzoff RK, Kron KA, Chase MW (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden*, **84**, 1–49.
- Soltis DE, Soltis PS, Zanis MJ (2002) Phylogeny of seed plants based on evidence from eight genes. *American Journal of Botany*, **89**, 1670–1681.
- Soltis DE, Soltis PS (2004) *Amborella* not a “basal angiosperm”? Not so fast. *American Journal of Botany*, **91**, 997–1001.
- Soltis PS, Soltis DE (2013) Angiosperm phylogeny: a framework for studies of genome evolution. *Plant Genome Diversity*, **2**, 1–11.
- Soltis PS, Soltis DE, Chase MW (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature*, **402**, 402–404.
- Steele PR, Guisinger-Bellian M, Linder CR, Jansen RK (2008) Phylogenetic utility of 141 low-copy nuclear regions in taxa at different taxonomic levels in two distantly related families of rosids. *Molecular Phylogenetics and Evolution*, **48**, 1013–1026.
- Stefanović S, Rice DW, Palmer JD (2004) Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? *BMC Evolutionary Biology*, **4**, 35.
- Stevens P (2001) Angiosperm phylogeny website. <http://www.mobot.org/MOBOT/research/APweb/>.
- Strand A, Leebens-Mack J, Milligan B (1997) Nuclear DNA-based markers for plant evolutionary biology. *Molecular Ecology*, **6**, 113–118.
- Sun G, Ji Q, Dilcher DL, Zheng SL, Nixon KC, Wang XF (2002) Archaeofractaceae, a new basal angiosperm family. *Science*, **296**, 899–904.
- Sun G, Dilcher DL, Wang HS, Chen ZD (2011) A eudicot from the Early Cretaceous of China. *Nature*, **471**, 625–628.
- Sun YX, Moore MJ, Meng AP, Soltis PS, Soltis DE, Li JQ, Wang HC (2013) Complete plastid genome sequencing of trochodendraceae reveals a significant expansion of the inverted repeat and suggests a paleogene divergence between the two extant species. *PLoS ONE*, **8**, e60429.
- Swamy B (1953) The morphology and relationships of the Chloranthaceae. *Journal of the Arnold Arboretum*, **34**, 375–408.
- Tang L, Zou XH, Achoundong G, Potgieter C, Second G, Zhang DY, Ge S (2010) Phylogeny and biogeography of the rice tribe (Oryzeae): evidence from combined analysis of 20 chloroplast fragments. *Molecular Phylogenetics and Evolution*, **54**, 266–277.
- Tank DC, Sang T (2001) Phylogenetic utility of the Glycerol-3-Phosphate acyltransferase gene: evolution and implications in *Paeonia* (Paeoniaceae). *Molecular Phylogenetics and Evolution*, **19**, 421–429.
- Thorne RF (1992) Classification and geography of the flowering plants. *The Botanical Review*, **58**, 225–327.
- Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. *Nature*, **418**, 671–677.
- Townsend JP, Su Z, Tekle YI (2012) Phylogenetic signal and noise: predicting the power of a data set to resolve phylogeny. *Systematic Biology*, **61**, 835–849.
- Wang HX, Moore MJ, Soltis PS, Bell CD, Brockington SF, Alexandre R, Davis CC, Latvis M, Manchester SR, Soltis DE (2009) Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences, USA*, **106**, 3853–3858.
- Wang X, Zhao L, Eaton D, Li D, Guo Z (2013) Identification of SNP markers for inferring phylogeny in temperate bamboos (Poaceae: Bambusoideae) using RAD sequencing. *Molecular Ecology Resources*, **13**, 938–945.
- Weiss CA, Garnaat CW, Mukai K, Hu Y, Ma H (1994) Isolation of cDNAs encoding guanine nucleotide-binding protein beta-subunit homologues from maize (*ZGB1*) and *Arabidopsis* (*AGB1*). *Proceedings of the National Academy of Sciences, USA*, **91**, 9554–9558.
- Wen J, Nie ZL, Soejima A, Meng Y (2007) Phylogeny of Vitaceae based on the nuclear *GAIL* gene sequences. *Canadian Journal of Botany*, **85**, 731–745.
- Wen J, Xiong Z, Nie ZL, Mao L, Zhu Y, Kan XZ, Ickert-Bond SM, Gerrath J, Zimmer EA, Fang XD (2013) Transcriptome sequences resolve deep relationships of the grape family. *PLoS ONE*, **8**, e74394.
- Westwood JH, Yoder JJ, Timko MP, dePamphilis CW (2010) The evolution of parasitism in plants. *Trends in Plant Science*, **15**, 227–235.
- Weng ML, Ruhlman TA, Gibby M, Jansen RK (2012) Phylogeny, rate variation, and genome size evolution of *Pelargonium* (Geraniaceae). *Molecular Phylogenetics and Evolution*, **64**, 654–670.
- Wilf P, Labandeira CC, Kress WJ, Staines CL, Windsor DM, Allen AL, Johnson KR (2000) Timing the radiations of leaf beetles: Hispines on gingers from latest Cretaceous to recent. *Science*, **289**, 291–294.

- Won H, Renner SS (2003) Horizontal gene transfer from flowering plants to Gnetum. *Proceedings of the National Academy of Sciences, USA*, **100**, 10824–10829.
- Worberg A, Quandt D, Barniske AM, Löhne C, Hilu KW, Borsch T (2007) Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organisms Diversity and Evolution*, **7**, 55–77.
- Wortley AH, Rudall PJ, Harris DJ, Scotland RW (2005) How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. *Systematic Biology*, **54**, 697–709.
- Wu FN, Mueller LA, Crouzillat D, Pétiard V, Tanksley SD (2006) Combining bioinformatics and phylogenetics to identify large sets of single-copy orthologous genes (*COSII*) for comparative, evolutionary and systematic studies: a test case in the euasterid plant clade. *Genetics*, **174**, 1407–1420.
- Wu ZQ, Ge S (2012) The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. *Molecular Phylogenetics and Evolution*, **62**, 573–578.
- Xi ZX, Ruhfel BR, Schaefer H, Amorim AM, Sugumaran M, Wurdack KJ, Endress PK, Matthews ML, Stevens PF, Mathews S (2012) Phylogenomics and a posteriori data partitioning resolve the Cretaceous angiosperm radiation Malpighiales. *Proceedings of the National Academy of Sciences, USA*, **109**, 17519–17524.
- Xu GX, Ma H, Nei M, Kong HZ (2009) Evolution of *F-box* genes in plants: different modes of sequence divergence and their relationships with functional diversification. *Proceedings of the National Academy of Sciences, USA*, **106**, 835–840.
- Yang ZH (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution*, **39**, 306–314.
- Yuan YW, Liu C, Marx HE, Olmstead RG (2009) The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *New Phytologist*, **182**, 272–283.
- Yuan YW, Liu C, Marx HE, Olmstead RG (2010) An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: phylogeny of Verbenaceae and the Verbenaceae complex. *Molecular Phylogenetics and Evolution*, **54**, 23–35.
- Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ (2002) The root of the angiosperms revisited. *Proceedings of the National Academy of Sciences, USA*, **99**, 6848–6853.
- Zhang N, Zeng LP, Shan HY, Ma H (2012) Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytologist*, **195**, 923–937.
- Zhang YJ, Ma PF, Li DZ (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). *PLoS ONE*, **6**, e20596.
- Zhu X, Chase MW, Qiu YL, Kong HZ, Dilcher DL, Li JH, Chen ZD (2007) Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids. *BMC Evolutionary Biology*, **7**, 217.
- Zimmer EA, Wen J (2013) Reprint of: using nuclear gene data for plant phylogenetics: progress and prospects. *Molecular Phylogenetics and Evolution*, **66**, 539–550.

(责任编辑: 王艇 责任编辑: 周玉荣)