博士毕业论文答辩

蔷薇类COM支深层次系统关系冲突探讨

—系统发育基因组学方法

答 辩 人：孙　苗

指导老师：陈之端 研究员

2014. 05. 28

报告内容

1 研究背景

2 研究方案

3 结果与讨论

4 结论与展望

报告内容

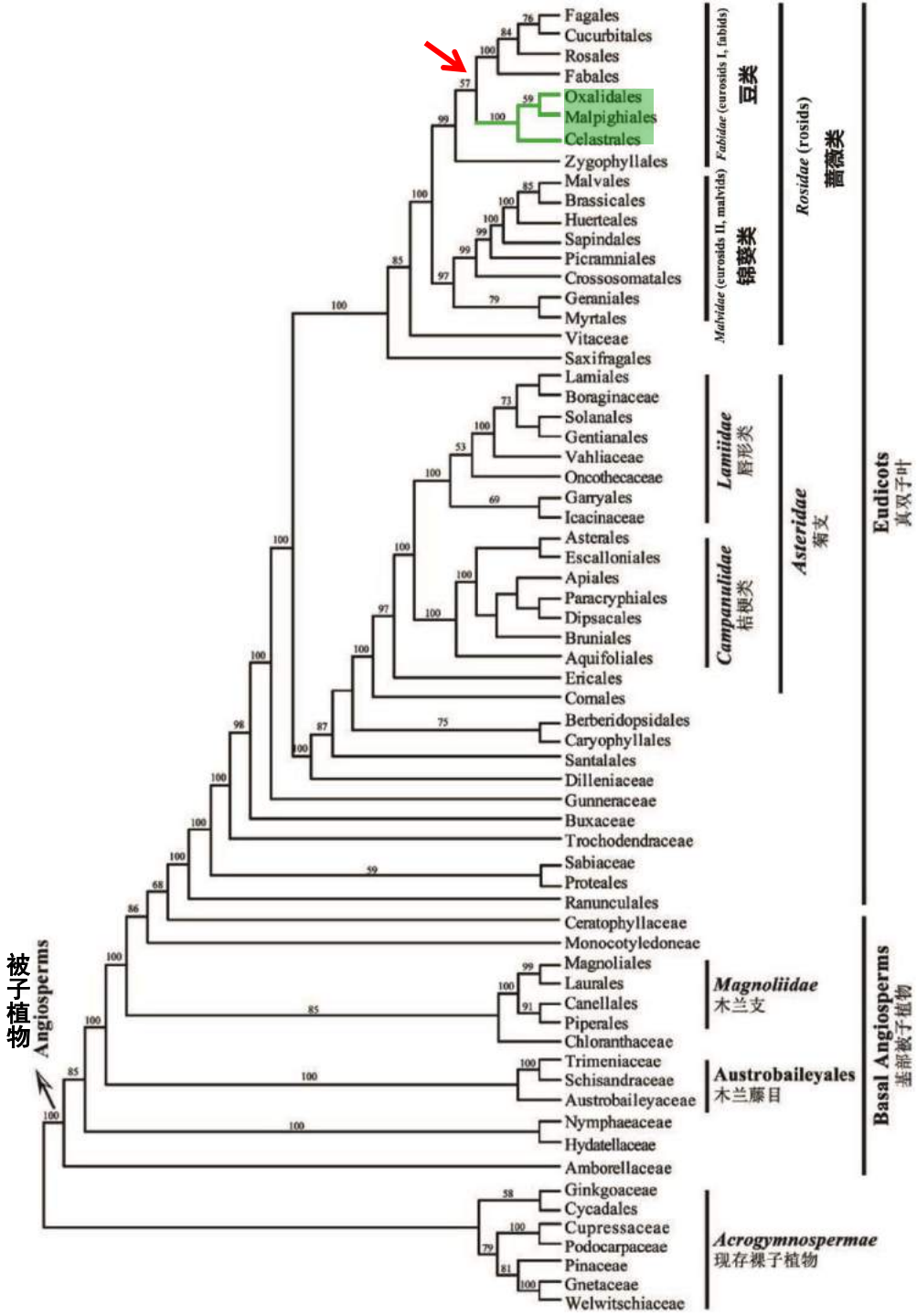
1 研究背景

2 研究方案

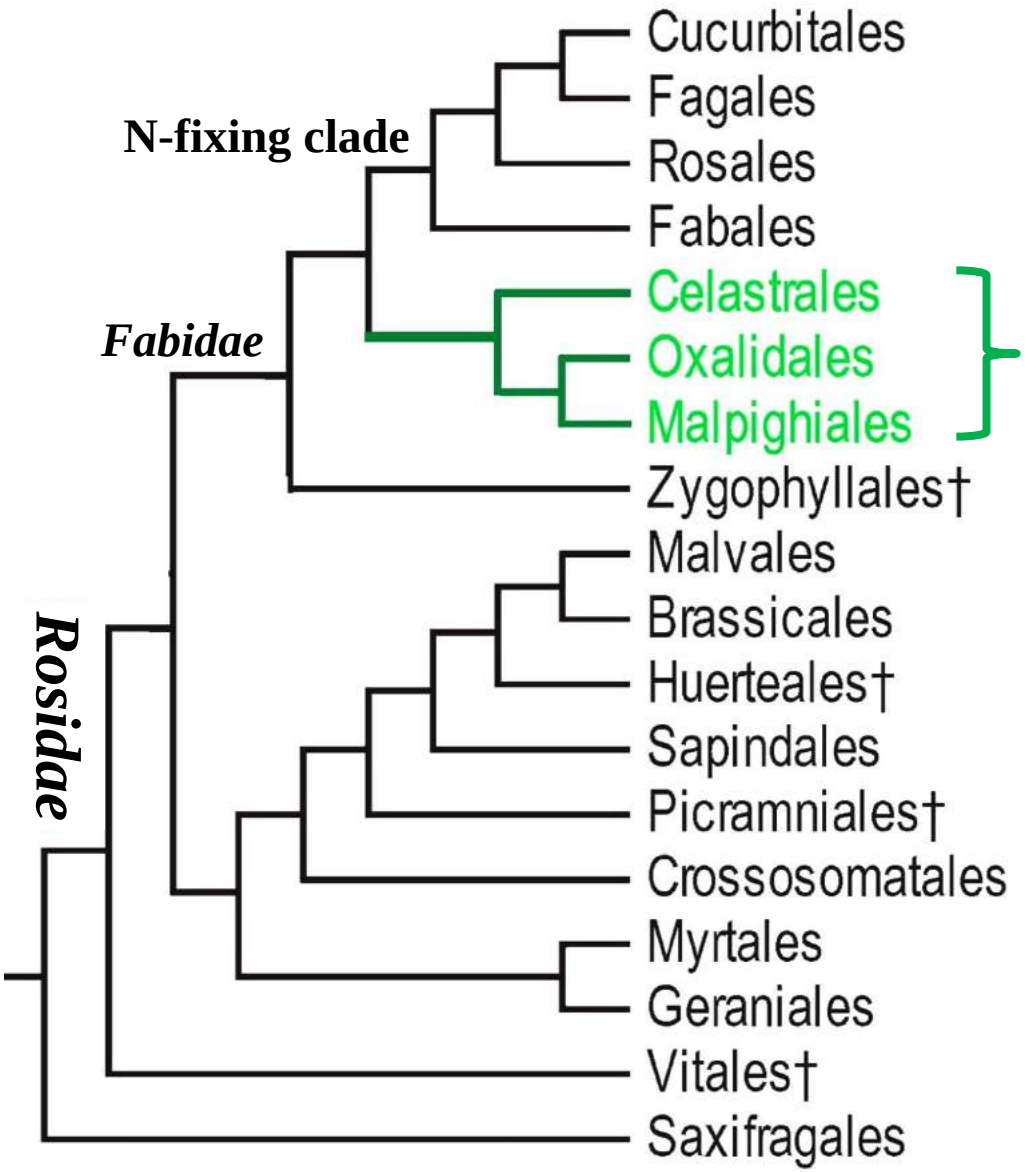
3 结果与讨论

4 结论与展望

研究背景


Fagales
Cucurbitales
Rosales
Fabales
Oxalidales
Malpighiales
Celastrales
Zygophyllales
Malvales
Brassicales
Huerteales
Sapindales
Picramniales
Crossosomatales
Geraniales
Myrtales
Vitaceae
Saxifragales
Lamiales
Boraginaceae
Solanales
Gentianales
Vahliaceae
Oncothecaceae
Garryales
Icacinaceae
Asterales
Escalloniales
Apiales
Paracryphiales
Dipsacales
Bruniales
Aquifoliales
Ericales
Cornales
Berberidopsidales
Caryophyllales
Santalales
Dilleniaceae
Gunneraceae
Buxaceae
Trochodendraceae
Sabiaceae
Proteales
Ranunculales
Ceratophyllaceae
Monocotyledoneae
Magnoliales
Laurales
Canellales
Piperales
Chloranthaceae
Trimeniaceae
Schisandraceae
Austrobaileyaceae
Nymphaeaceae
Hydatellaceae
Amborellaceae
Ginkgoaceae
Cycadales
Cupressaceae
Podocarpaceae
Pinaceae
Gnetaceae
Welwitschiaceae
Fabidae (eurosids I, fabids)
豆类
Malvidae (eurosids II, malvids)
锦葵类
Rosidae (rosids)
蔷薇类
Lamiidae
唇形类
Campanulidae
桔梗类
Asteridae
菊支
Eudicots
真双子叶
Magnoliidae
木兰支
Austrobaileyales
木兰藤目
Basal Angiosperms
基部被子植物
Acrogymnospermae
现存裸子植物
被子植物 Angiosperms

COM支概念



COM支 (COM clade)

范围：

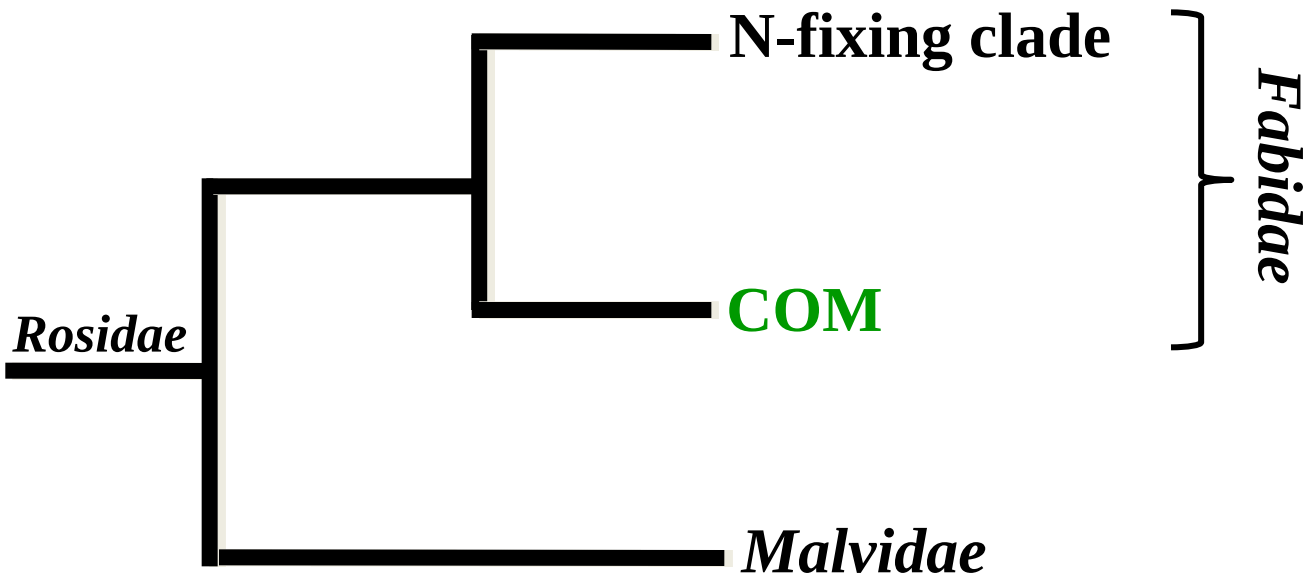
- ~1/3 蔷薇类
- 48-50 科
- 870 属
- 19105 种

Endress & Matthew, 2006
APG III, 2009

叶绿体基因研究

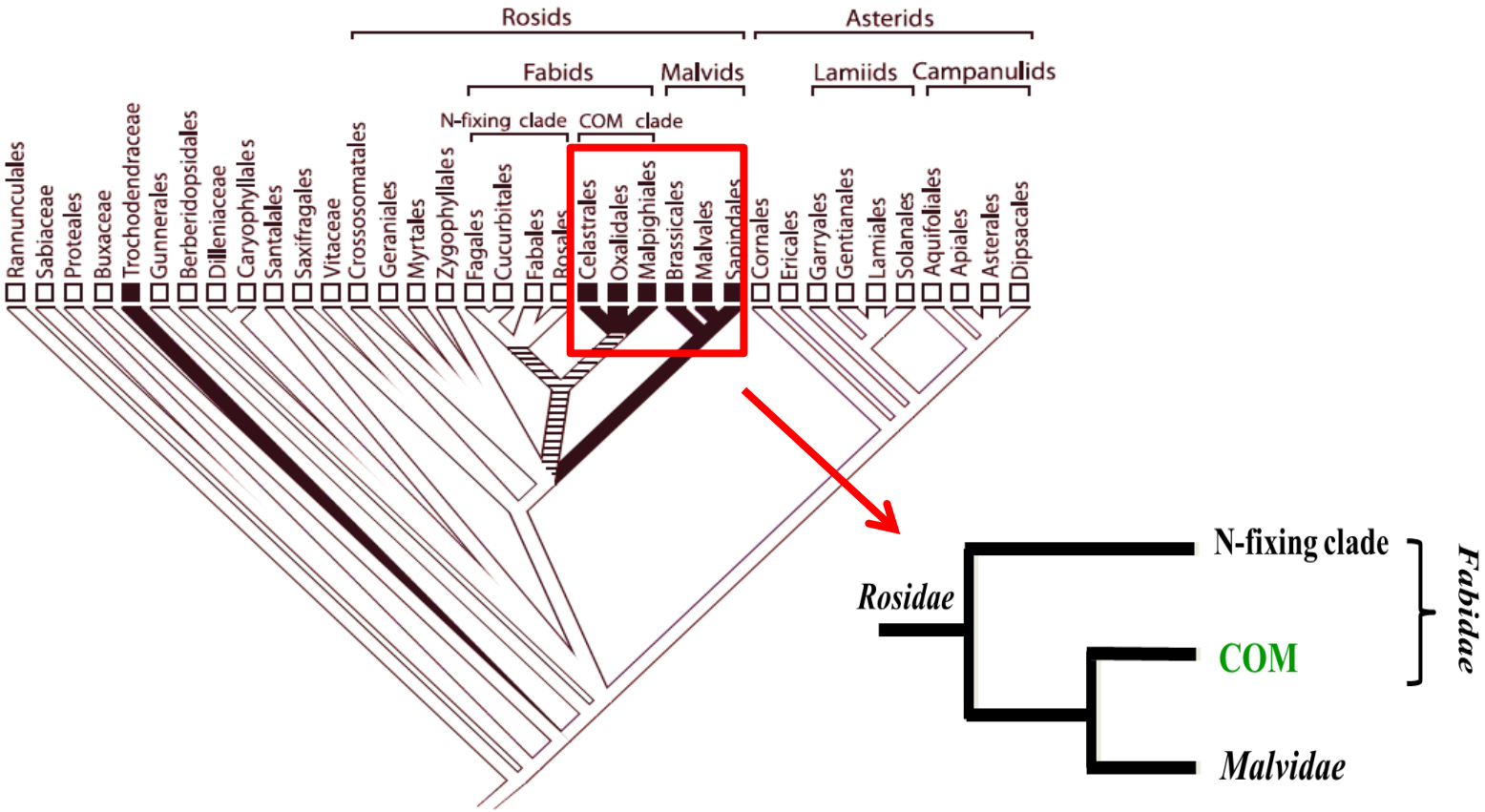
	系统位置	分析方法/支持率	分子标记	类群总取样	COM支取样	参考文献
叶绿体基因	Nr	Character-state weighting/–	*rbcL*	499	–	Chase et al. 1993
	COM + *Fabidae*	Parsimony/52% JK; BI/1.0 PP	*matK*	374	16	Hilu et al. 2003
	COM + *Fabidae*	Parsimony jackknifing/77% JK; BI/1.0 PP	*rbcL*, *atpB*, 18S rDNA	560	64	Soltis et al. 2000; Soltis et al.2007
	COM + *Fabidae*	ML/100% BS; MP/79% BS; BI/1.0 PP	81 cp	64	3	Jansen et al. 2007
	COM + *Fabidae*	ML/89% BS	*rbcL*, *atpB*, m*atK*, 18S rDNA, 26S rDNA	567	59	Burleigh et al. 2009
	COM + *Fabidae*	ML/100% BS	10 cp, 2 nu	117	33	Wang et al. 2009
	COM + *Fabidae*	ML/53% BS	83 cp	86	5	Moore et al. 2010
	COM + *Fabidae*	ML/99% BS	IR	244	14	Moore et al. 2011
	COM + *Fabidae*	ML/57% BS	11 cp, 2 nu, 4 mt	640	154	Soltis et al. 2011
	COM + *Fabida*	ML/81% BS, 70% BS, 82% BS, 69% BS (ntAll, ntNo3rd, RY, AA)	78 cp	360	9	Ruhfel et al. 2014

叶绿体基因研究



基于叶绿体基因的研究一致支持COM支与*Fabidae*近缘

形态学研究



COM支与*Malvidae*花结构形态**共近裔特征：受精时珠被的内层比外层更厚**

Endress & Matthew, 2006

线粒体基因研究

BMC Evolutionary Biology

BioMed Central

Research article

Open Access

Mitochondrial *matR* sequences help to resolve deep phylogenetic relationships in rosids

Xin-Yu Zhu[1,2], Mark W Chase[3], Yin-Long Qiu[4], Hong-Zhi Kong[1], David L Dilcher[5], Jian-Hua Li[6] and Zhi-Duan Chen*[1]

Abstract

Background: Rosids are a major clade in the angiosperms containing 13 orders and about one-third of angiosperm species. Recent molecular analyses recognized two major groups (i.e., fabids with seven orders and malvids with three orders). However, phylogenetic relationships within the two groups and among fabids, malvids, and potentially basal rosids including Geraniales, Myrtales, and Crossosomatales remain to be resolved with more data and a broader taxon sampling. In this study, we obtained DNA sequences of the mitochondrial *matR* gene from 174 species representing 72 families of putative rosids and examined phylogenetic relationships and phylogenetic utility of *matR* in rosids. We also inferred phylogenetic relationships within the "rosid clade" based on a combined data set of 91 taxa and four genes including *matR*, two plastid genes (*rbcL*, *atpB*), and one nuclear gene (18S rDNA).

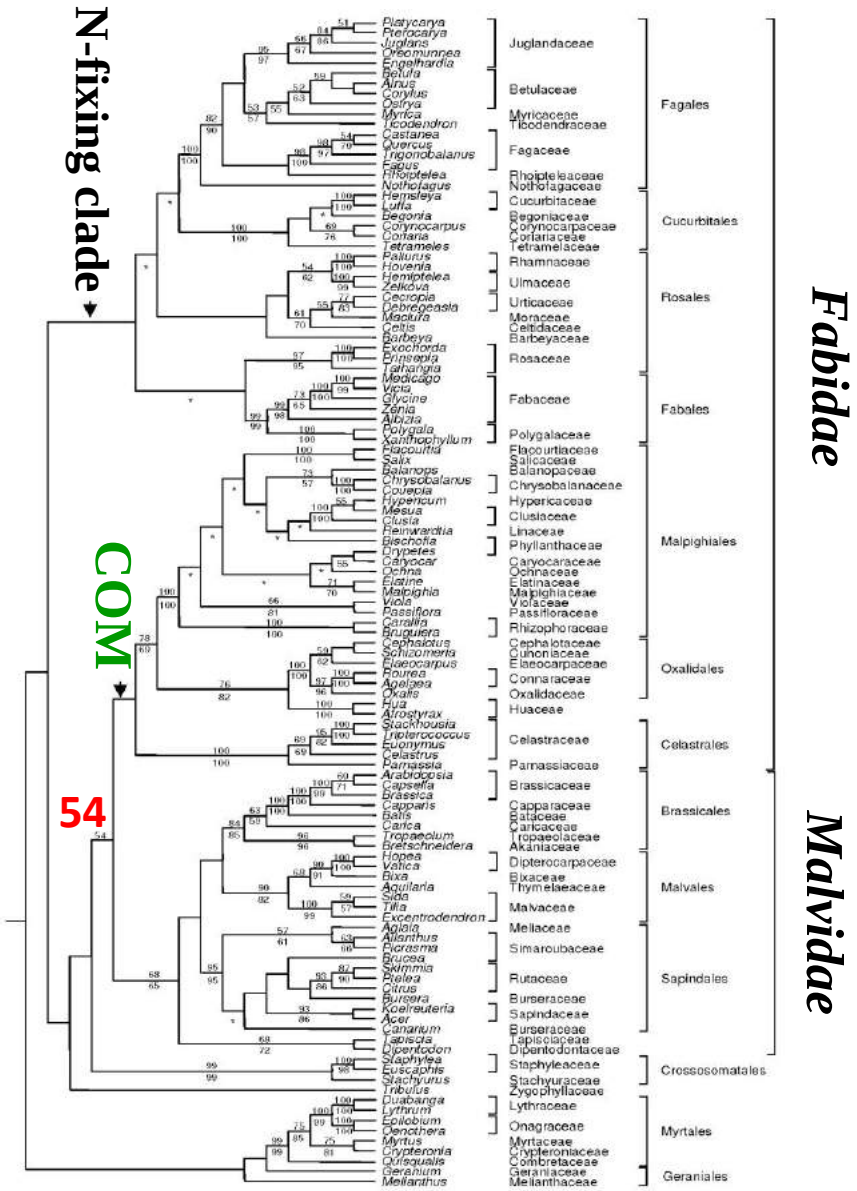
Results: Comparison of mitochondrial *matR* and two plastid genes (*rbcL* and *atpB*) showed that the synonymous substitution rate in *matR* was approximately four times slower than those of *rbcL* and *atpB*; however, the nonsynonymous substitution rate in *matR* was relatively high, close to its synonymous substitution rate, indicating that the *matR* has experienced a relaxed evolutionary history. Analyses of our *matR* sequences supported the monophyly of malvids and most orders of the rosids. However, fabids did not form a clade; instead, the COM clade of fabids (Celastrales, Oxalidales, Malpighiales, and Huaceae) was sister to malvids. Analyses of the four-gene data set suggested that Geraniales and Myrtales were successively sister to other rosids, and that Crossosomatales were sister to malvids.

Conclusion: Compared to plastid genes such as *rbcL* and *atpB*, slowly evolving *matR* produced less homoplasious but not less informative substitutions. Thus, *matR* appears useful in higher-level angiosperm phylogenetics. Analysis of *matR* alone identified a novel deep relationship within rosids, the grouping of the COM clade of fabids and malvids, which was not resolved by any previous molecular analyses but recently suggested by floral structural features. Our four-gene analysis supported the placements of Geraniales, Myrtales at basal nodes of the rosid clade and placed Crossosomatales as sister to malvids. We also suggest that the core part of rosids should include fabids, malvids and Crossosomatales.

Zhu et al., 2007

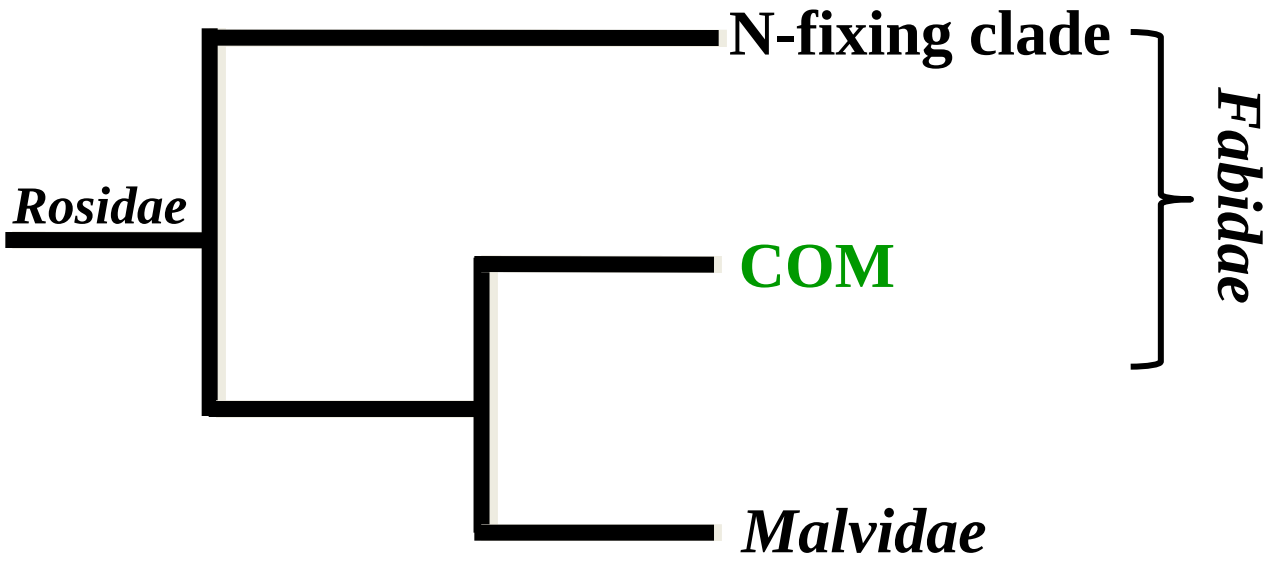
线粒体基因研究

- 支持COM支与*Malvidae*近缘
- 与形态一致
- 分子系统学证据



Zhu et al., 2007

线粒体基因研究



Zhu et al., 2007

线粒体基因研究

JSE Journal of Systematics and Evolution

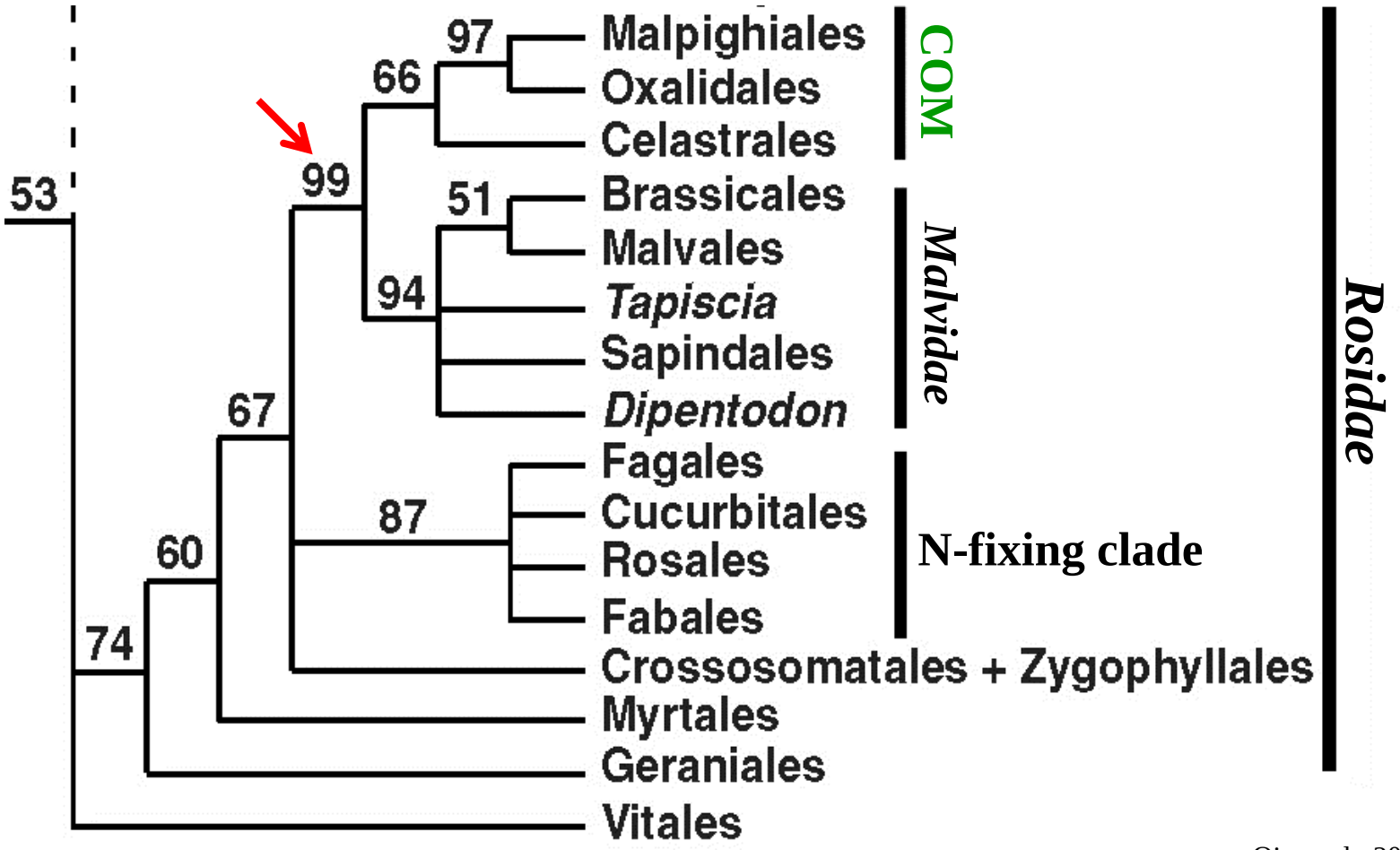
Angiosperm phylogeny inferred from sequences of four mitochondrial genes

[1]Yin-Long QIU* [1]Libo LI [1]Bin WANG [1,2]Jia-Yu XUE [1]Tory A. HENDRY [1]Rui-Qi LI
[1]Joseph W. BROWN [1]Yang LIU [1]Geordan T. HUDSON [3]Zhi-Duan CHEN

Abstract An angiosperm phylogeny was reconstructed in a maximum likelihood analysis of sequences of four mitochondrial genes, *atp1, matR, nad5*, and *rps3*, from 380 species that represent 376 genera and 296 families of seed plants. It is largely congruent with the phylogeny of angiosperms reconstructed from chloroplast genes *atpB, matK,* and *rbcL*, and nuclear 18S rDNA. The basalmost lineage consists of *Amborella* and Nymphaeales (including Hydatellaceae). Austrobaileyales follow this clade and are sister to the mesangiosperms, which include Chloranthaceae, *Ceratophyllum*, magnoliids, monocots, and eudicots. With the exception of Chloranthaceae being sister to *Ceratophyllum,* relationships among these five lineages are not well supported. In eudicots, Ranunculales, Sabiales, Proteales, Trochodendrales, Buxales, Gunnerales, Saxifragales, Vitales, Berberidopsidales, and Dilleniales form a basal grade of lines that diverged before the diversification of rosids and asterids. Within rosids, the COM (Celastrales–Oxalidales–Malpighiales) clade is sister to malvids (or rosid II), instead of to the nitrogen-fixing clade as found in all previous large-scale molecular analyses of angiosperms. Santalales and Caryophyllales are members of an expanded asterid clade. This study shows that the mitochondrial genes are informative markers for resolving relationships among genera, families, or higher rank taxa across angiosperms. The low substitution rates and low homoplasy levels of the mitochondrial genes relative to the chloroplast genes, as found in this study, make them particularly useful for reconstructing ancient phylogenetic relationships. A mitochondrial gene-based angiosperm phylogeny provides an independent and essential reference for comparison with hypotheses of angiosperm phylogeny based on chloroplast genes, nuclear genes, and non-molecular data to reconstruct the underlying organismal phylogeny.

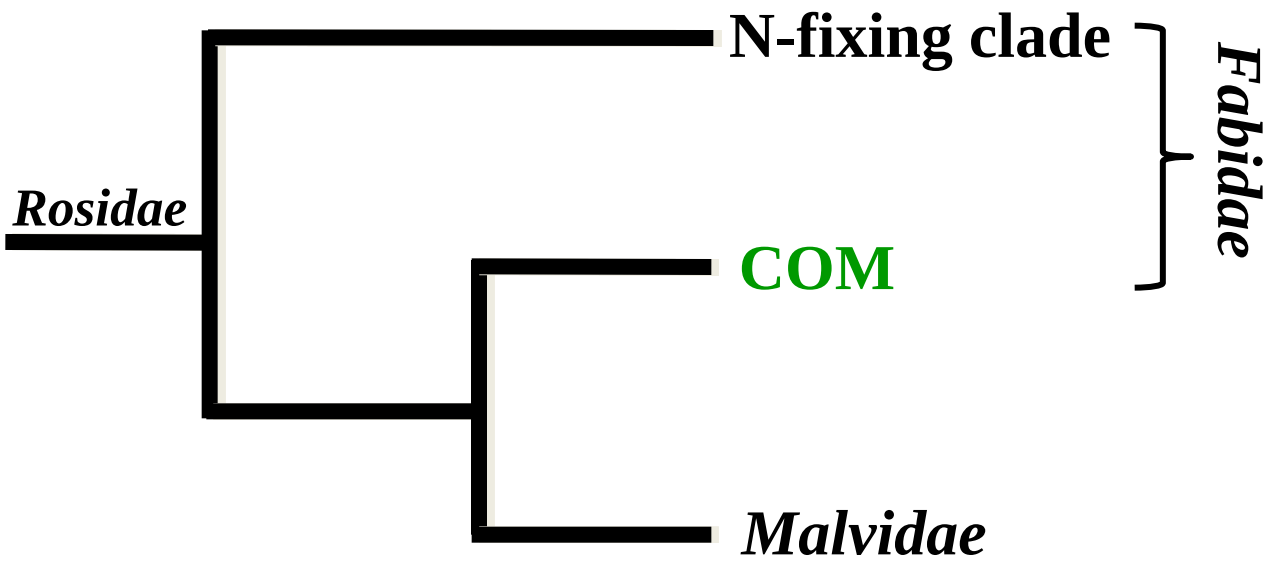
Qiu et al., 2010

线粒体基因研究



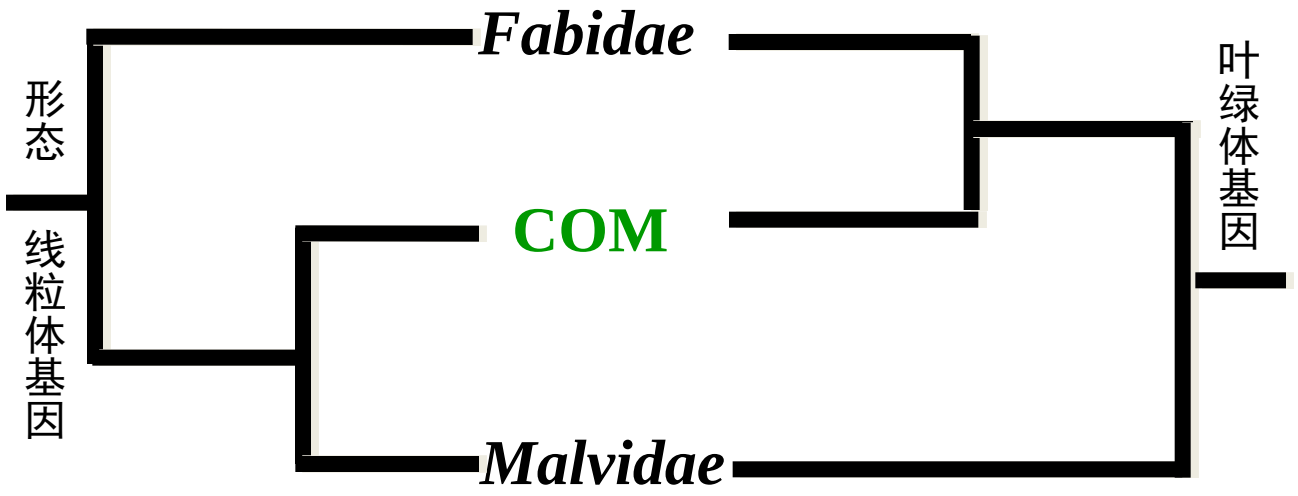
Qiu et al., 2010

线粒体基因研究



Qiu et al., 2010

分歧



- 核基因，你怎么看？

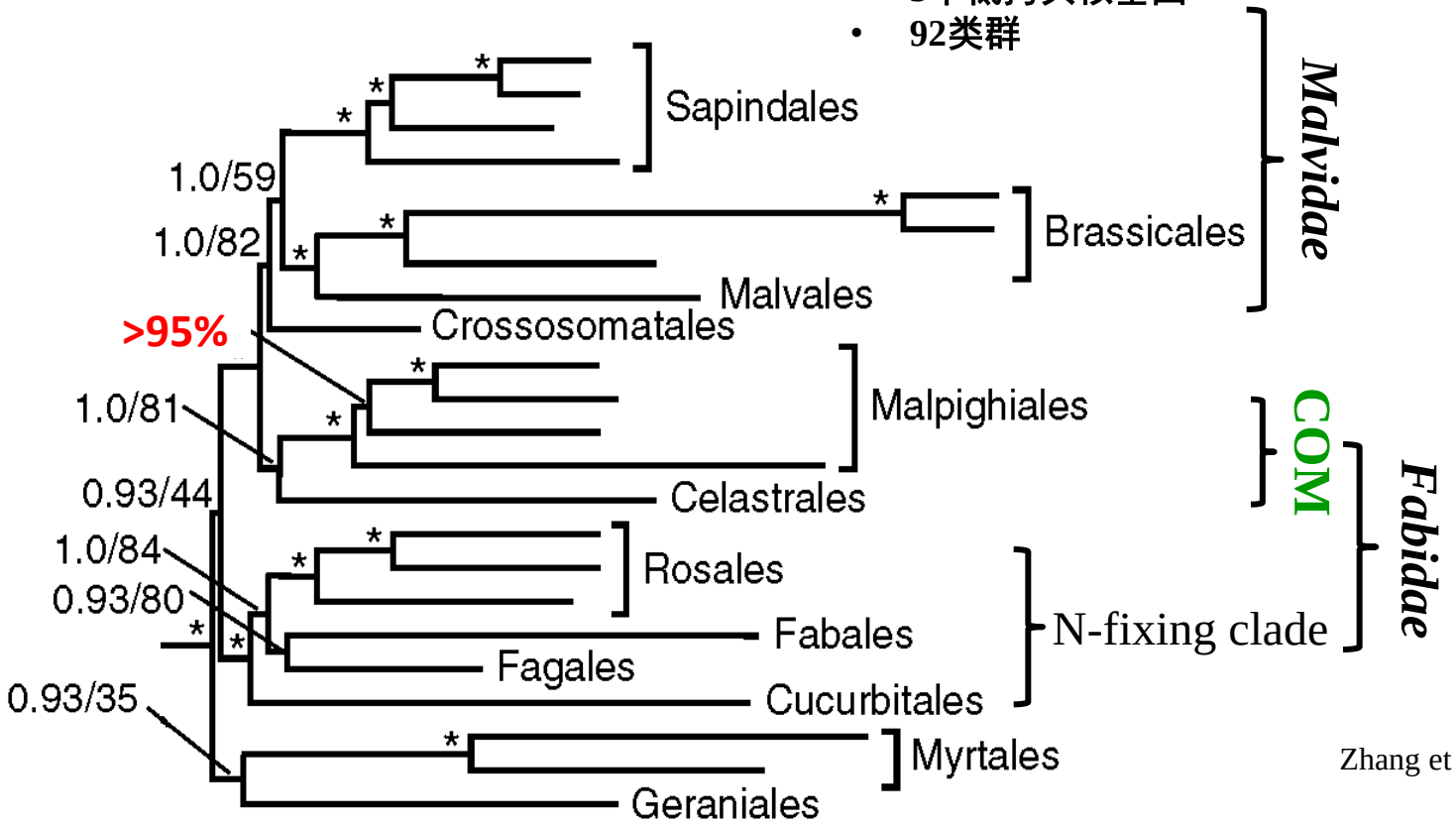
核基因研究

New Phytologist
Research

Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms

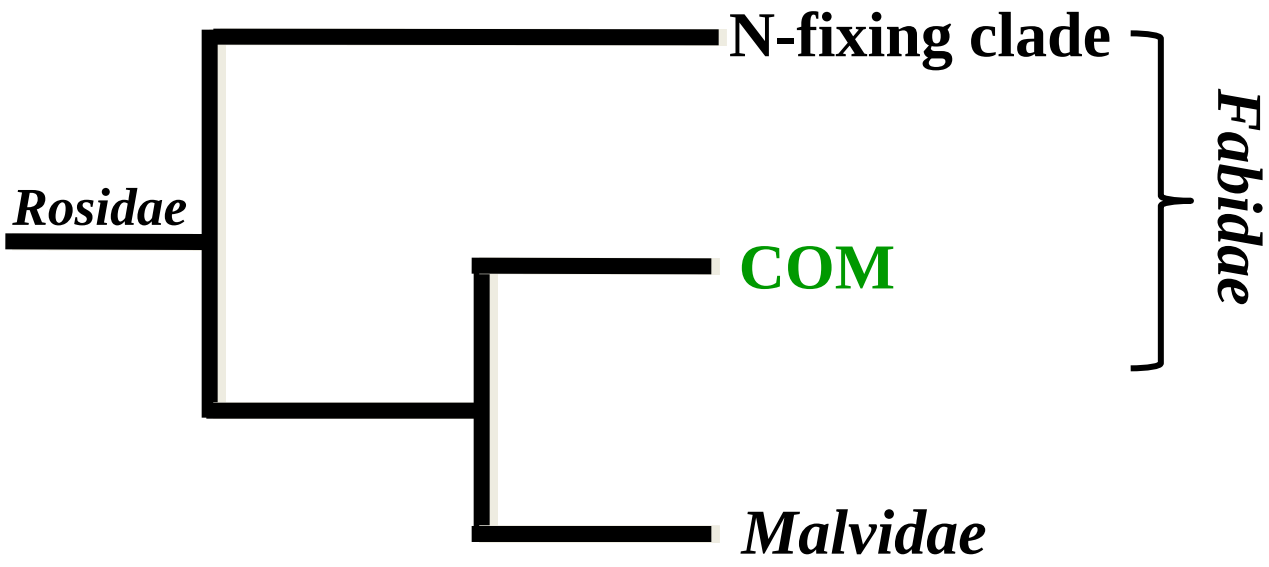
Ning Zhang[1], Liping Zeng[1], Hongyan Shan[2] and Hong Ma[1,3]

- 5个低拷贝核基因
- 92类群

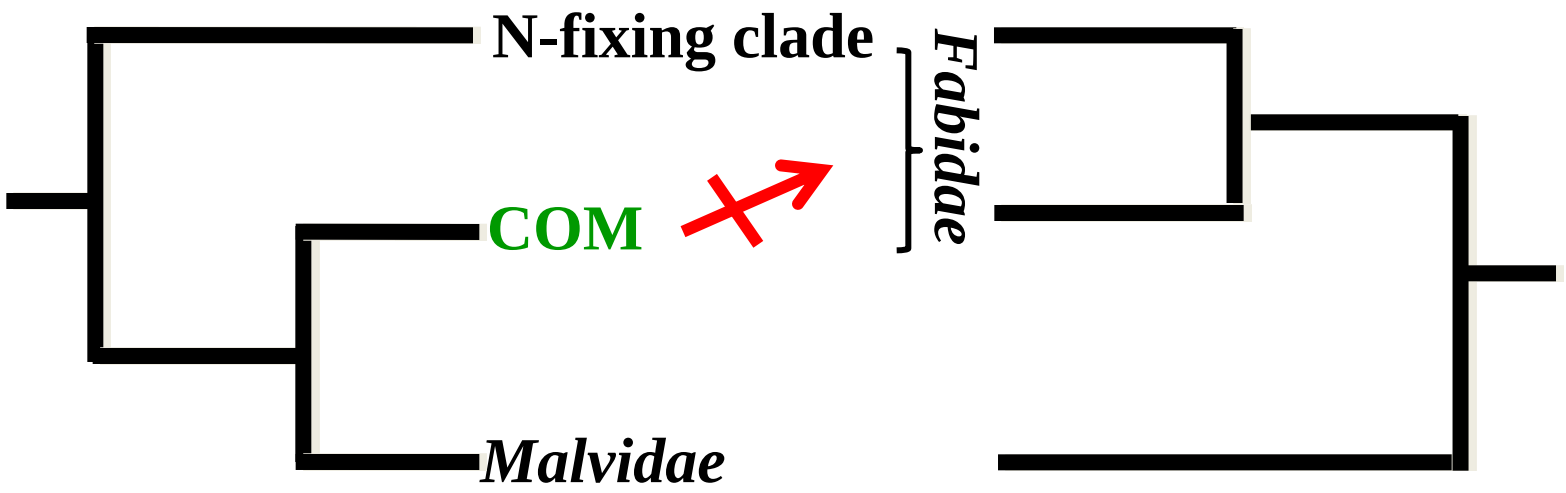


Zhang et al., 2012

核基因研究



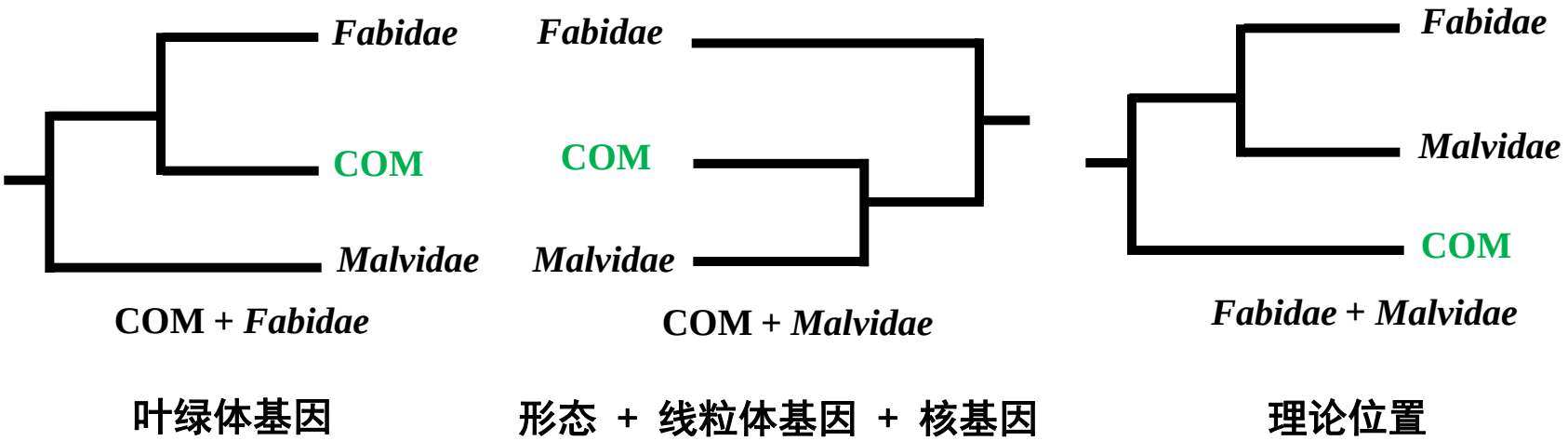
类群概念变化



蔷薇类内的**COM支、*Fabidae* (N-fixing clade)、*Malvidae***

分别作为三个不同的类群来对待

冲突的系统位置

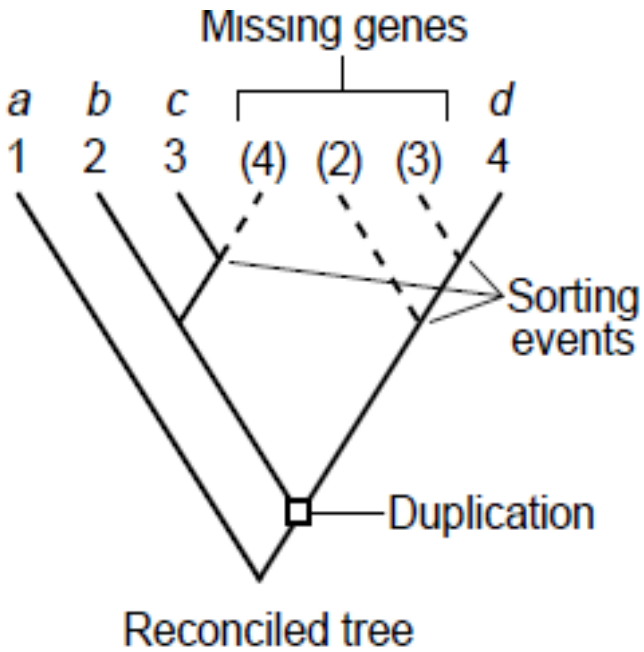


- 形态、线粒体、核基因及形态研究结果都支持COM支与*Malvidae*近缘，而叶绿体基因支持COM支和*Fabidae*更近缘，二者的结论相冲突
- 前人研究的取样策略、建树方法、以及选用基因各不相同

系统发育关系冲突概要

冲突原因	原因阐述		解决方案
软冲突	人为因素	数据不足	下一代测序技术 系统发育基因组学 取样的代表性与合理密度 基因选择 合理的进化模型
		取样偏差	
		基因选择不当	
		测序错误	
	序列因素	碱基替代模式偏差	第三密码子排除 **氨基酸序列建树** **RY编码** **快速进化位点移除** 一致网络分析法
		长枝吸引	
		进化速率异质性	
		进化饱和	
硬冲突	**生物过程**	快速辐射分化	最小遗传距离法 融合法 基因树简约法 网状进化网络分析法
		杂交/渐渗	
		不完全谱系筛选	
		基因水平转移	
		旁系同源	
		基因重复与丢失	
		基因重组	

不完全谱系筛选（incomplete lineage sorting，deep coalescence）

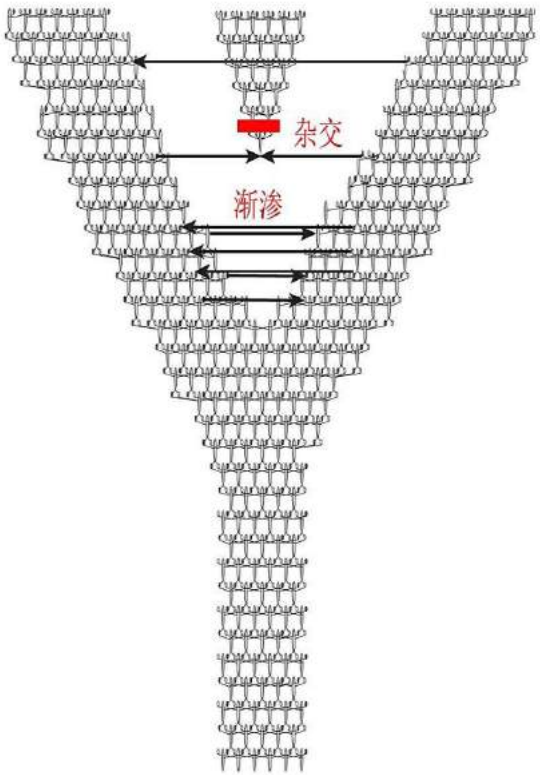


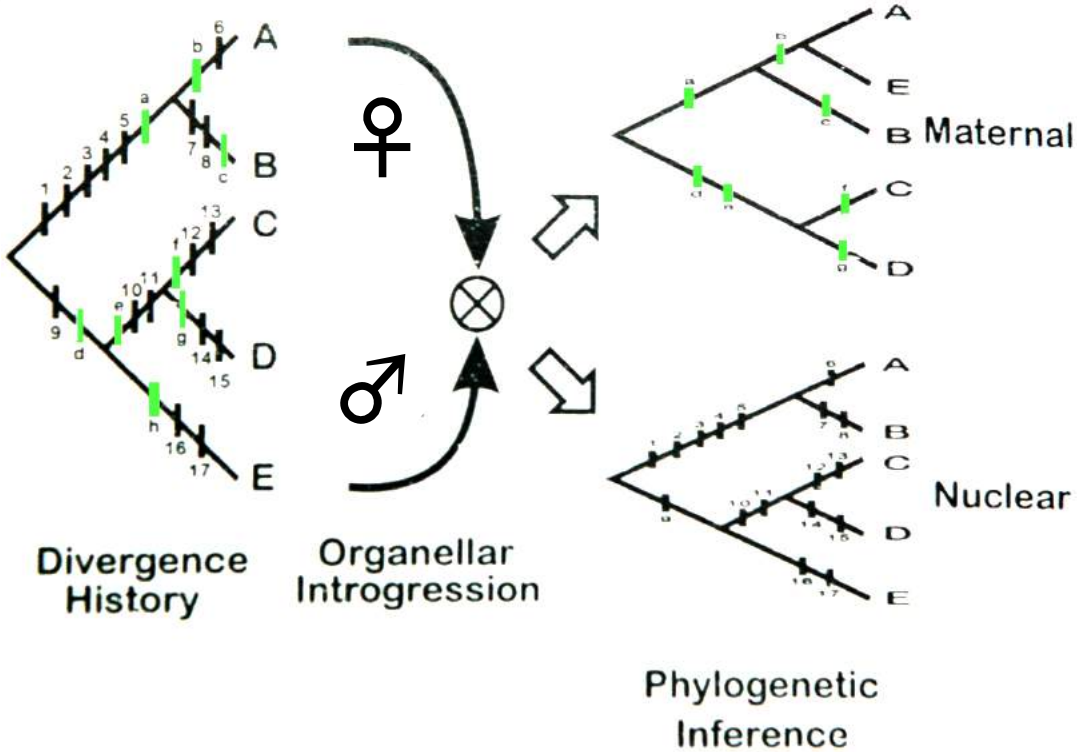
Page & Charleston 1998

- 祖先→基因（多态）→后代
- 自然选择或遗传漂变→不完全谱系筛选
- 近期的连续物种分化事件
- 有效种群
- 早期的分化影响更深次

Pamilo & Nei, 1988
Enard & Paabo, 2004
Degnan & Rosenberg, 2009

杂交/渐渗（Hybridization/introgression）





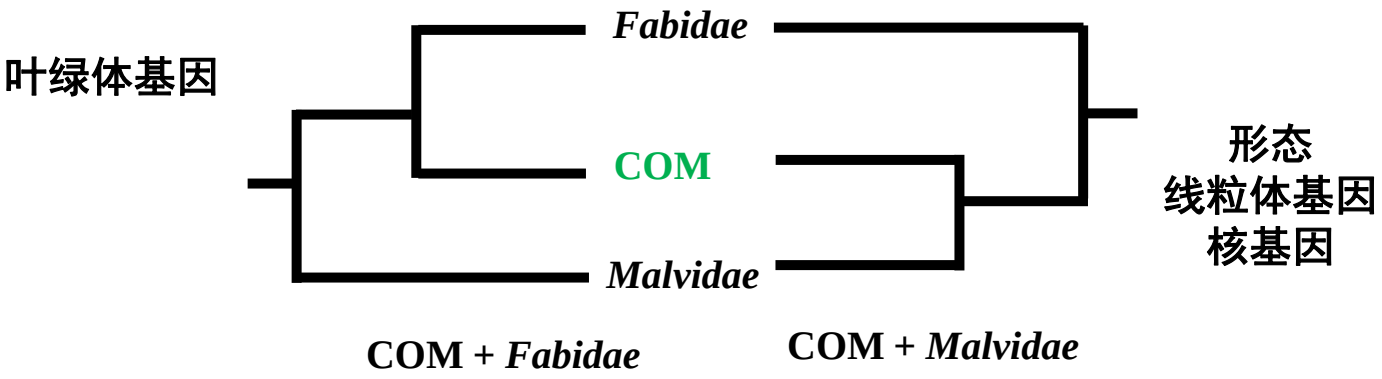
Simpson 2012

Wendell & Doyle, 1998

科学问题

COM支系统位置的冲突是否真实存在

如果存在，前人研究的取样策略、建树方法及选用的基因不尽相同，那么这种冲突是来自取样、序列、建树方法中的错误系统误差（软冲突），还是来自生物进化事件（硬冲突）



研究意义

- 纠正现行APG系统框架内重要分支的系统位置
- 解决蔷薇类一级分支上的重大分歧，为构建蔷薇类可靠的系统发育框架奠定基础
- 对理解蔷薇类，以至整个被子植物的系统发育与进化都是极为关键的
- 为验证、解决其他类群系统发育关系冲突的问题提供借鉴和参考

报告内容

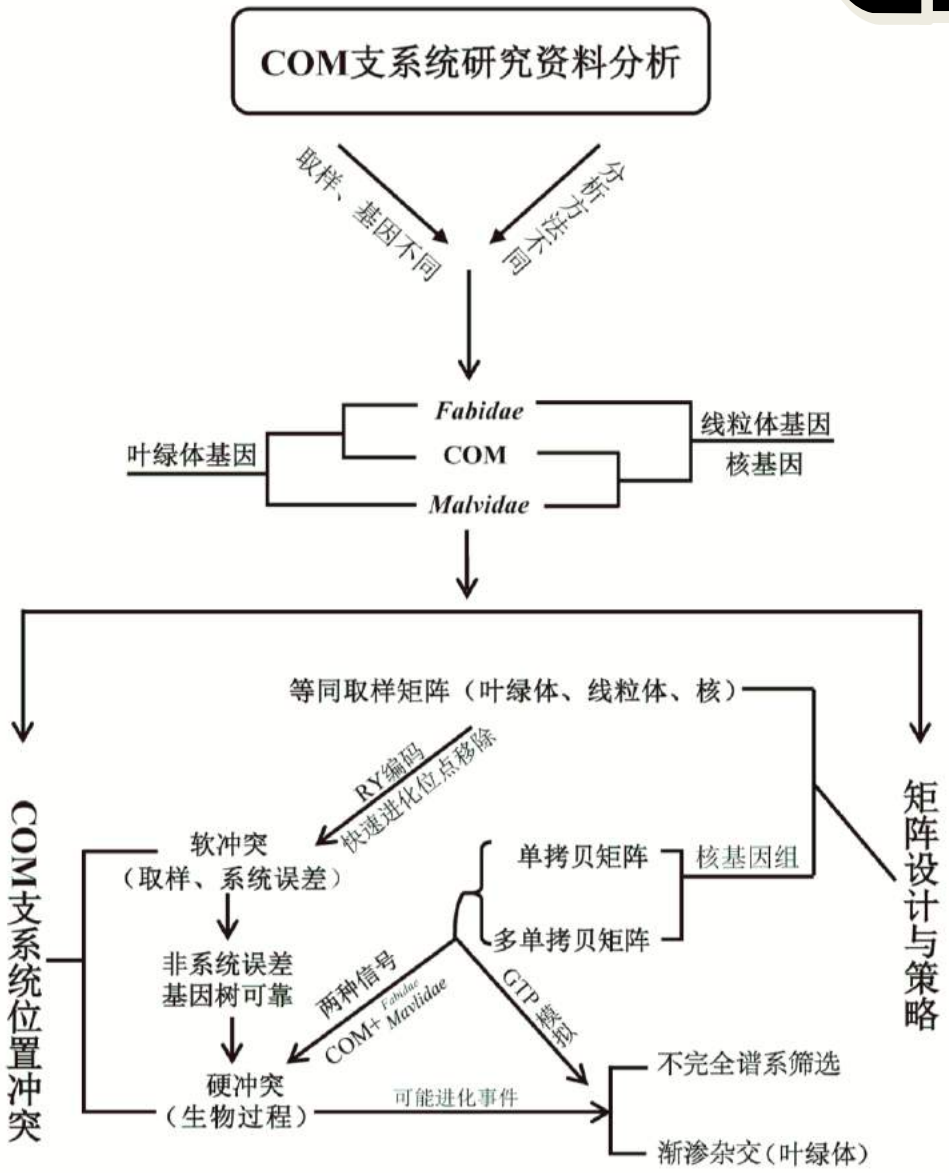
1 研究背景

2 研究方案

3 结果与讨论

4 结论与展望

研究方案


COM支系统研究资料分析
取样、基因不同
分析方法不同
Fabidae
COM
Malvidae
叶绿体基因
线粒体基因
核基因
等同取样矩阵（叶绿体、线粒体、核）
RY编码
快速进化位点移除
软冲突
（取样、系统误差）
非系统误差
基因树可靠
硬冲突
（生物过程）
两种信号
COM+ Fabidae Malvidae
单拷贝矩阵
多单拷贝矩阵
核基因组
GTP 数据
可能进化事件
不完全谱系筛选
渐渗杂交（叶绿体）
COM支系统位置冲突
矩阵设计与策略

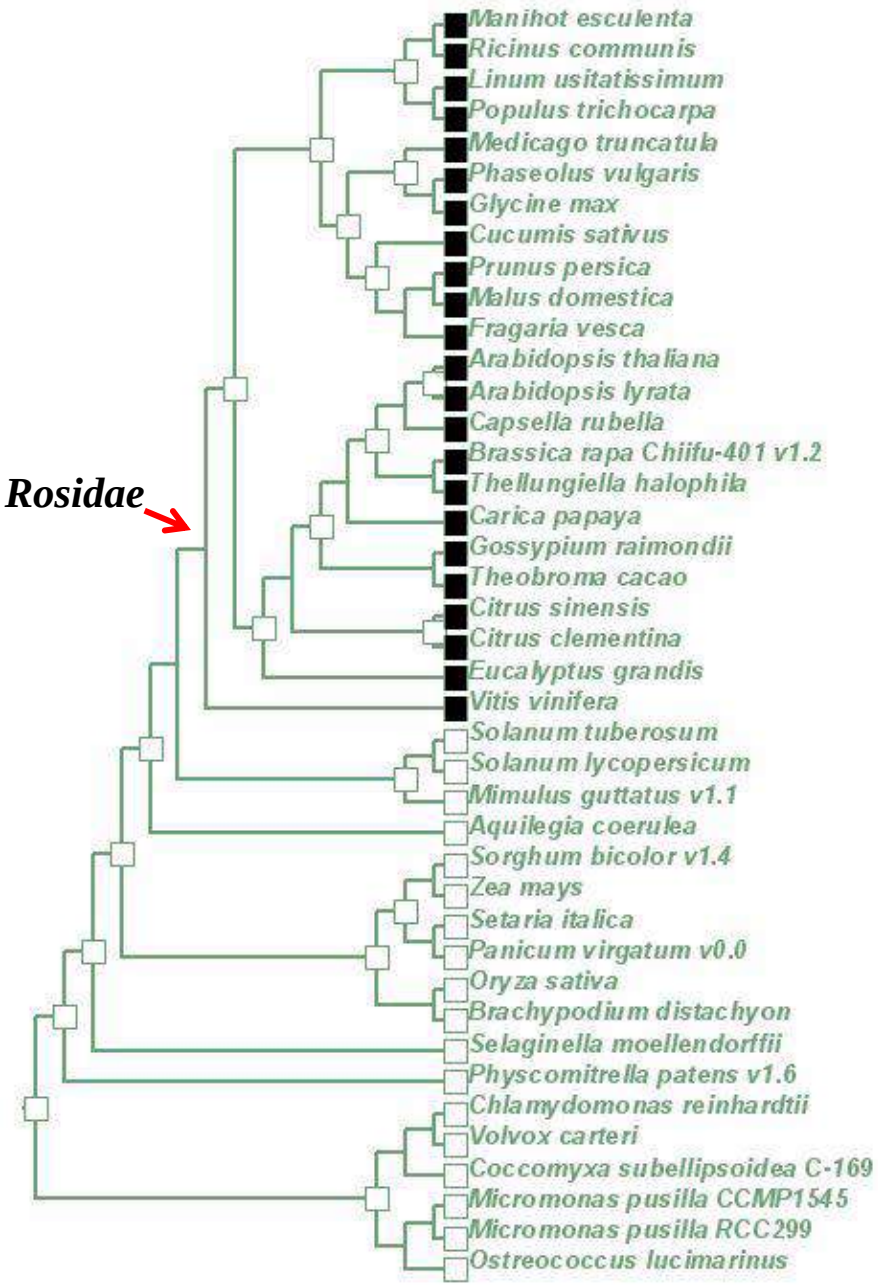
矩阵设计

一、验证COM支的系统发育位置

分别来自叶绿体、线粒体、核基因的，取样近等同的矩阵

二、寻找冲突根源

研究方案



- 41+7 种陆地植物已测全基因组
- 23种来自蔷薇类

http://www.phytozome.net/

矩阵设计

一、验证COM支的系统发育位置

分别来自叶绿体、线粒体、核基因的，取样近等同的矩阵

二、寻找冲突根源

大容量的单拷贝和多拷贝的核基因组数据

矩阵

矩阵类型	类群	基因	参考文献
叶绿体	82	78基因	Ruhfel et al. 2014
线粒体	79	*atp1*， *matR*， *nad5*， *rps3*	Qiu et al. 2010
核*	92	*SMC1* ，*SMC2*，*MCM5*，*MSH1*，*MLH1*	Zhang et al. 2012
核基因组	101	8,445个单拷贝直系同源核基因	Lee et al. 2011
核基因组	22	3,748个多拷贝核基因	Chen et al. 2006

兼顾各大分支的代表性，三个矩阵之间的目、科、属最大限度地同步匹配；各分支及其下类群的名称和范围遵守Cantino et al. （2007）和 APG III （2009）

对所有矩阵的最大似然法（Maximum Likelihood, ML）分析都是应用RAxML在GTRCAT模型下进行的

一、验证COM支的系统发育位置

核苷酸（单基因）

挖掘基本的系统发育信息

氨基酸

排除核苷酸转换和颠换的影响

降低转换饱和，进化速率及组成成份异质性

RY 编码

缓解长枝吸引和模型错配

快速位点移除

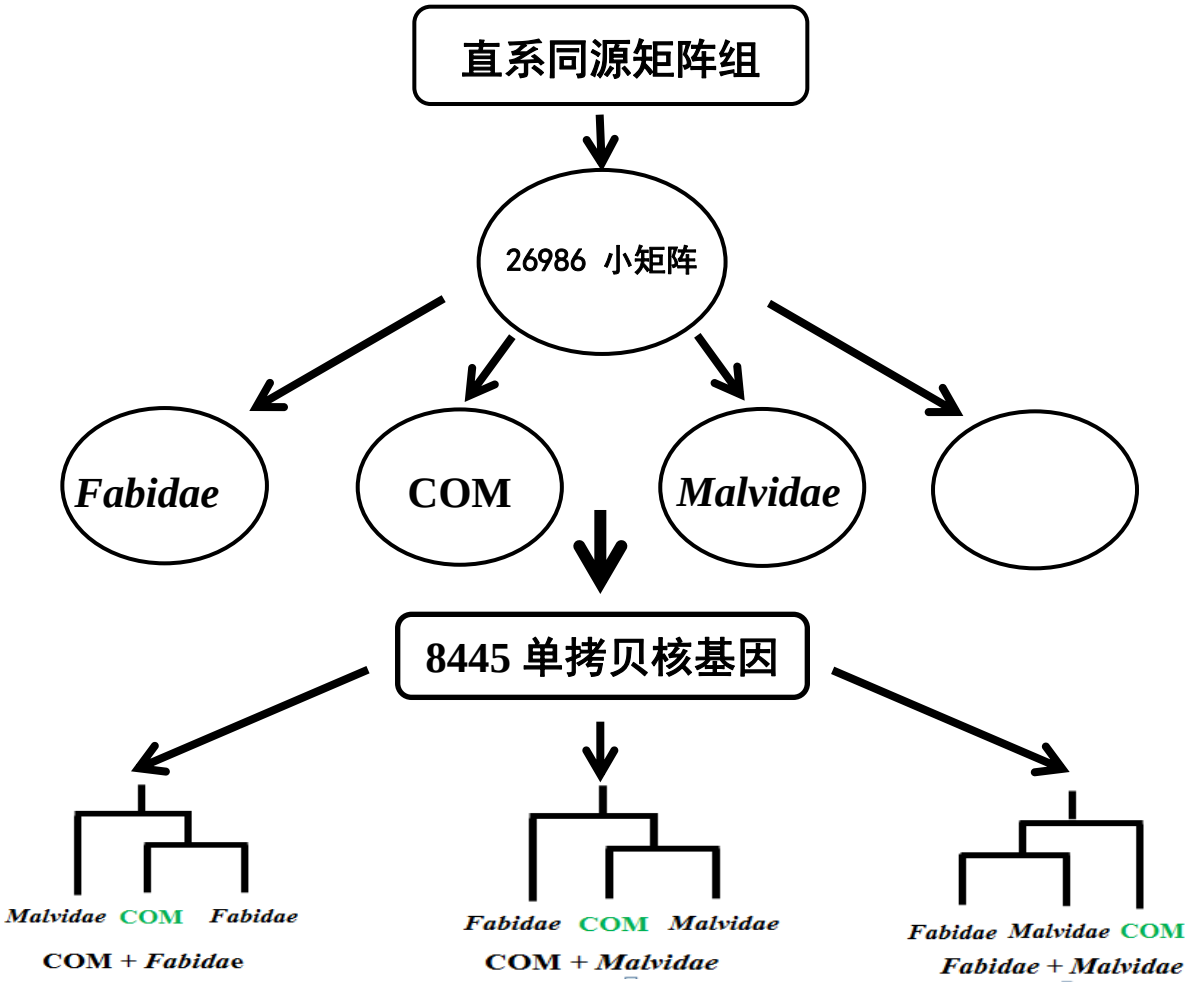
验证系统发育关系是否可靠

Hashimoto et al., 1995
Phillips & Penny, 2003
Harrison et al., 2004
Delsuc et al., 2005
Gibson et al., 2005
Philippe et al., 2005
Goremykin et al., 2010

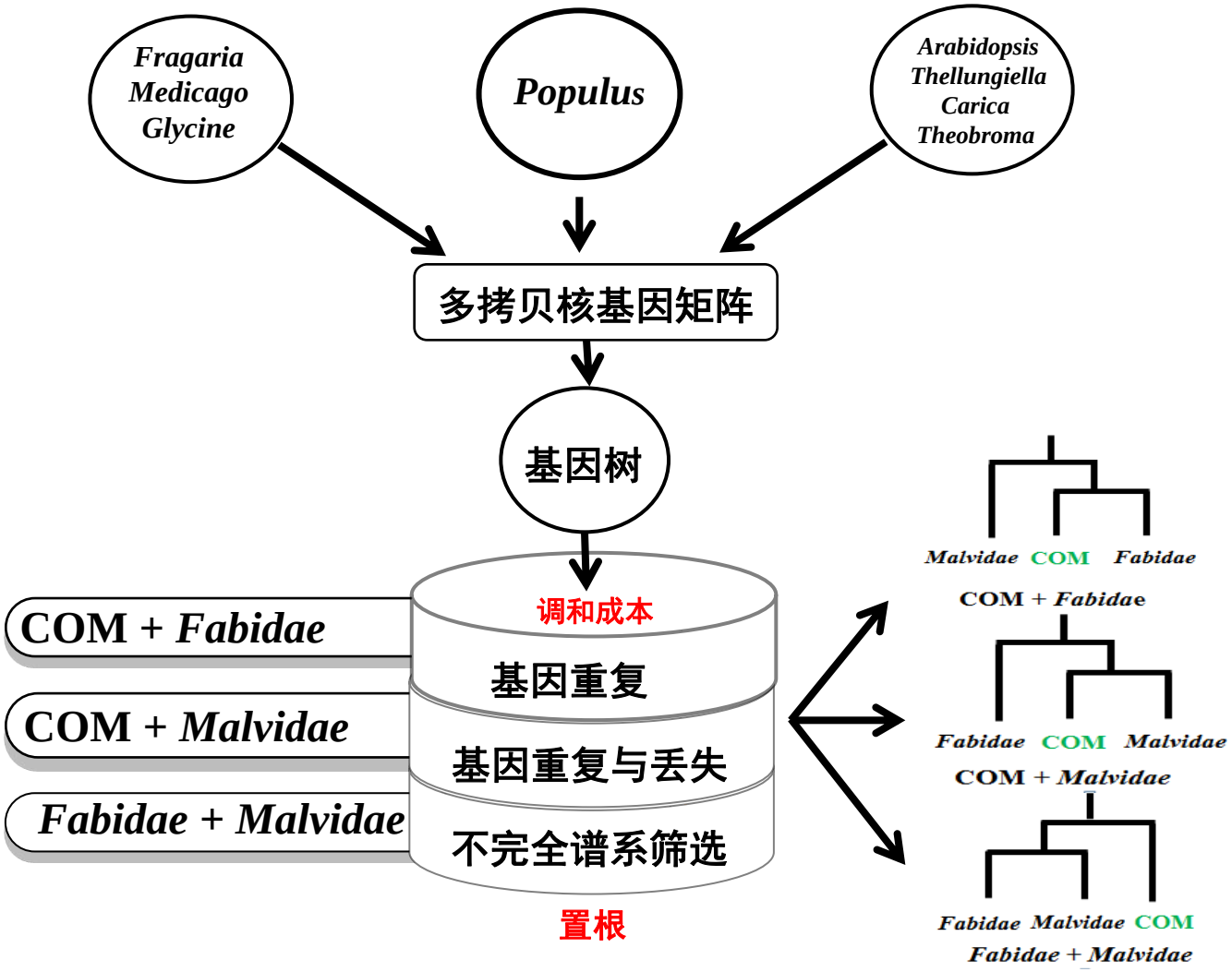
二、寻找冲突根源

大容量的单拷贝和多拷贝的核基因组数据

单拷贝核基因数据分析


直系同源矩阵组
26986 小矩阵
Fabidae
COM
Malvidae
8445 单拷贝核基因
Malvidae COM Fabidae
COM + Fabidae
Fabidae COM Malvidae
COM + Malvidae
Fabidae Malvidae COM
Fabidae + Malvidae

多拷贝核基因数据分析


Fragaria
Medicago
Glycine
Populus
Arabidopsis
Thellungiella
Carica
Theobroma
多拷贝核基因矩阵
基因树
调和成本
基因重复
基因重复与丢失
不完全谱系筛选
置根
COM + Fabidae
COM + Malvidae
Fabidae + Malvidae
Malvidae COM Fabidae
COM + Fabidae
Fabidae COM Malvidae
COM + Malvidae
Fabidae Malvidae COM
Fabidae + Malvidae

报告内容

1 研究背景

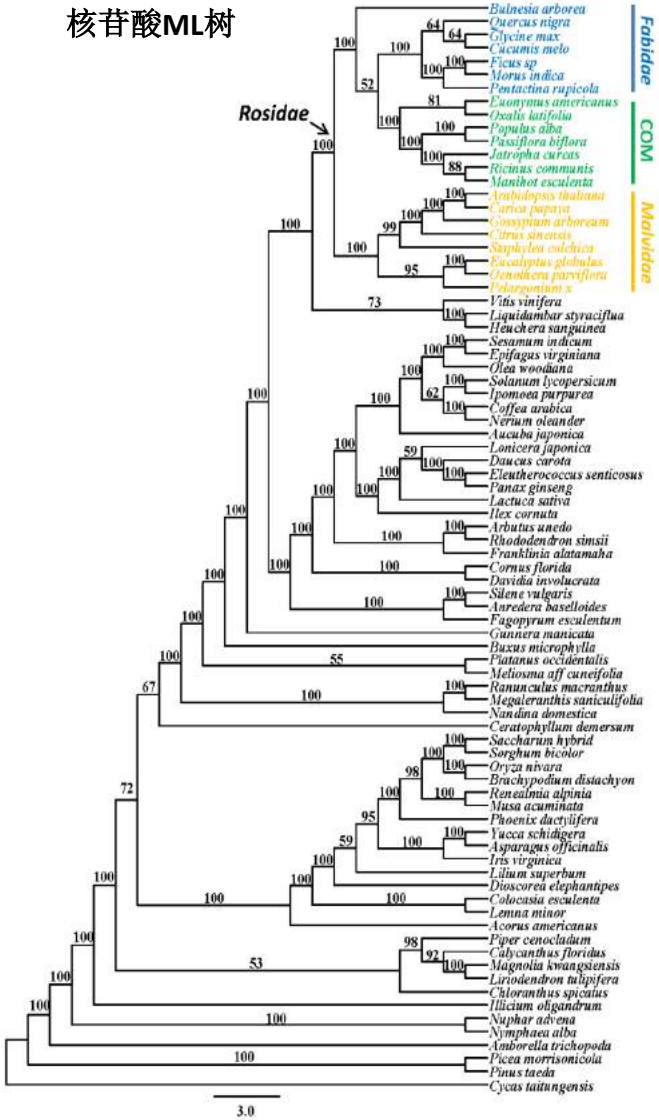
2 研究方案

3 结果与讨论

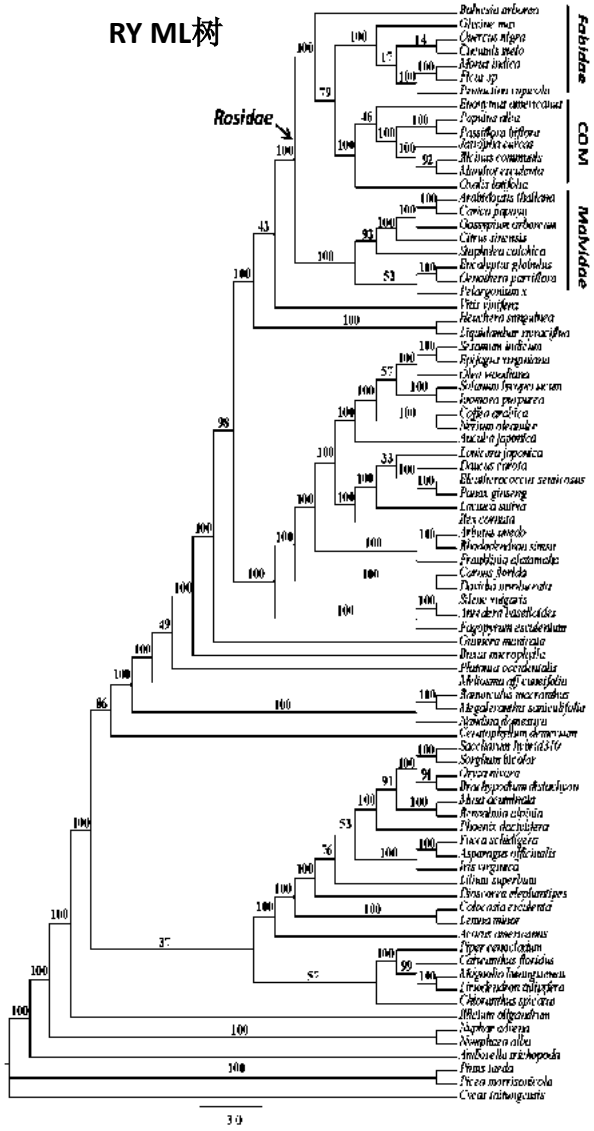
4 结论与展望

叶绿体基因分析结果

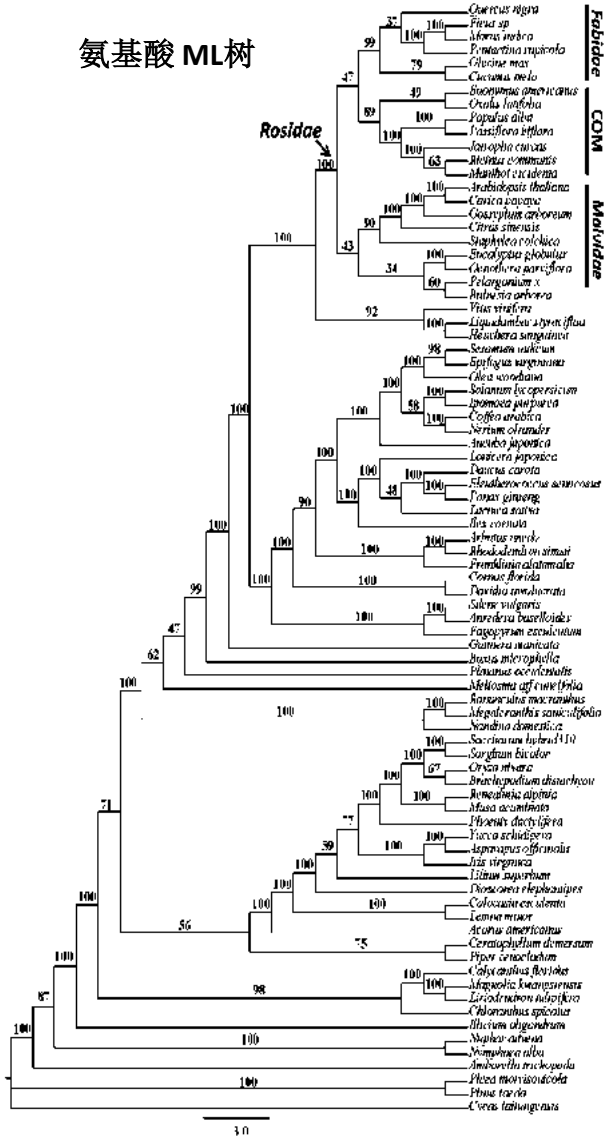
核苷酸ML树



RY ML树

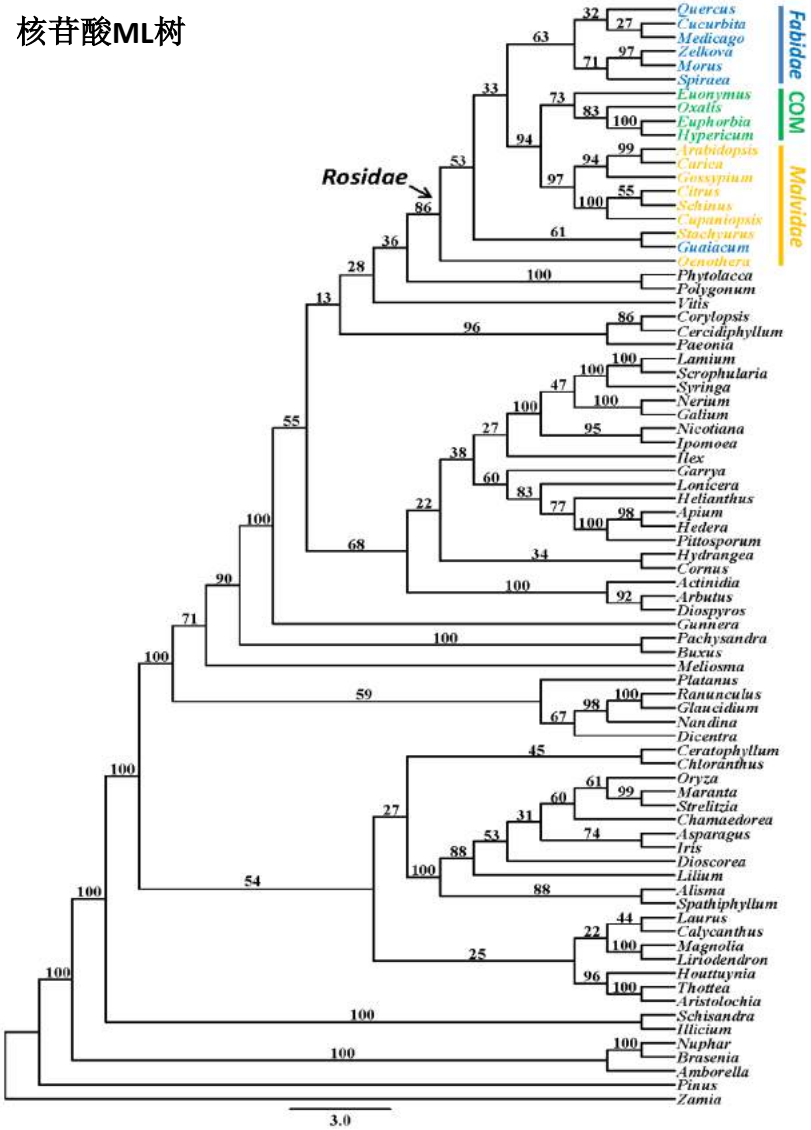


氨基酸 ML树

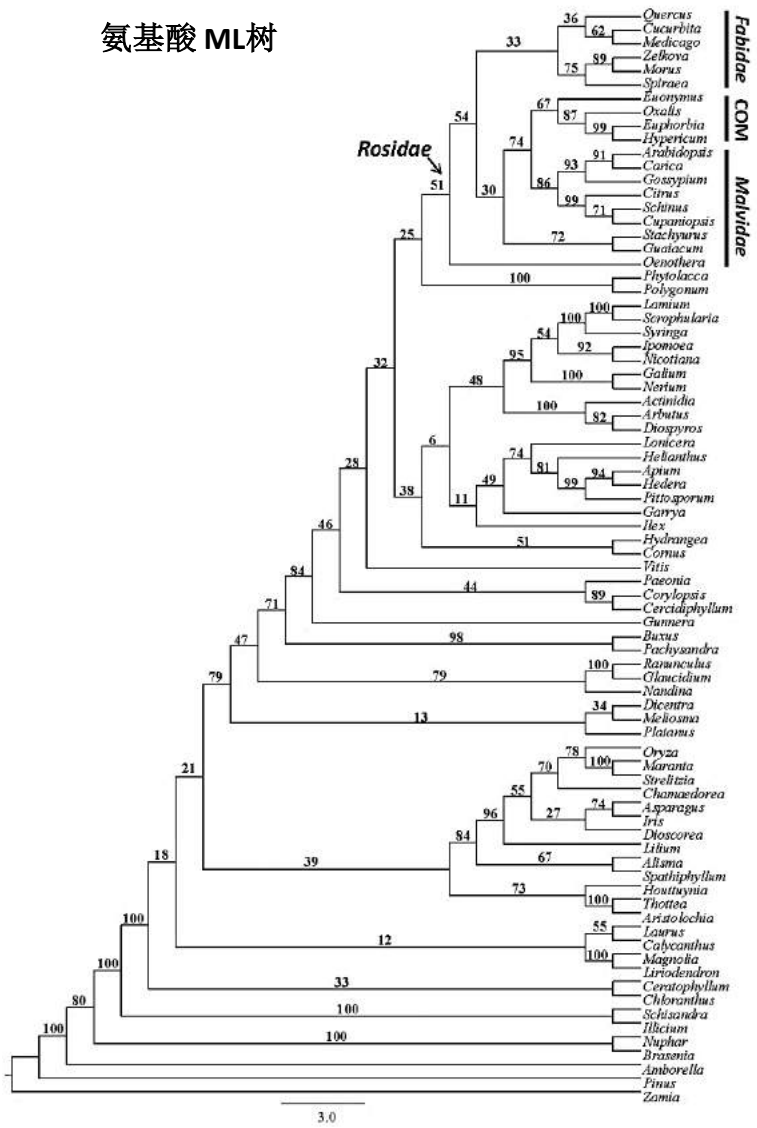


线粒体基因分析结果

核苷酸ML树

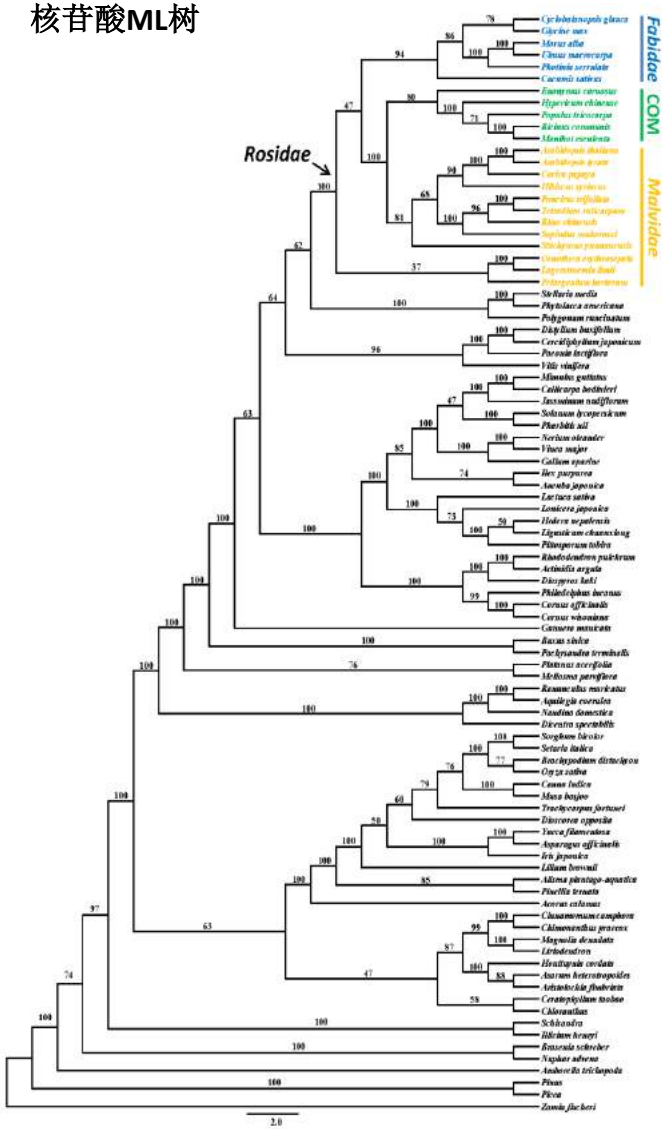


氨基酸 ML树

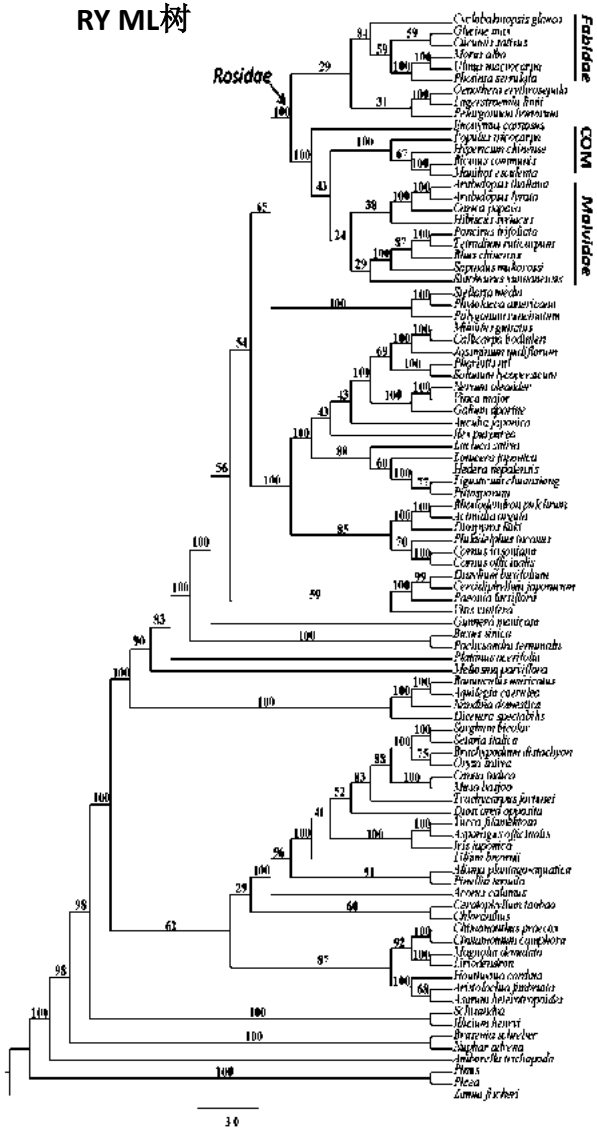


核基因分析结果

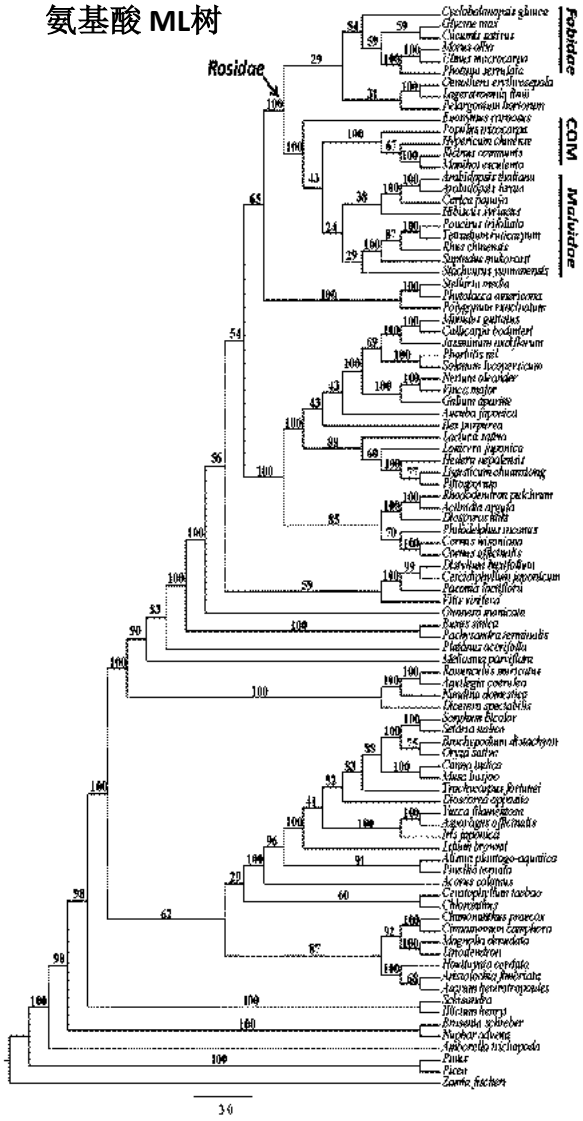
核苷酸ML树



RY ML树



氨基酸 ML树



系统发育关系验证

分析方法/矩阵	叶绿体矩阵 82-类群 78-基因	线粒体矩阵 79-类群 4-基因	核矩阵 92-类群 5-基因
核苷酸矩阵	100% BS（52%）	94% BS	100% BS
氨基酸矩阵	47% BS	74% BS	100% BS
RY编码矩阵	79% BS	——	100% BS
快速进化位点移除矩阵	5% → 98% BS 10% → 48% BS 20% → 1% BS	5% → 100% BS 10%→ 23% BS	5% → 100% BS 10% → 100% BS 20% → 96% BS
支持的拓扑结构	COM + *Fabidae*	COM + *Malvidae*	COM + *Malvidae*

各矩阵得到的系统发育关系是可靠的，并未受到系统误差和取样偏差的影响

系统发育关系验证

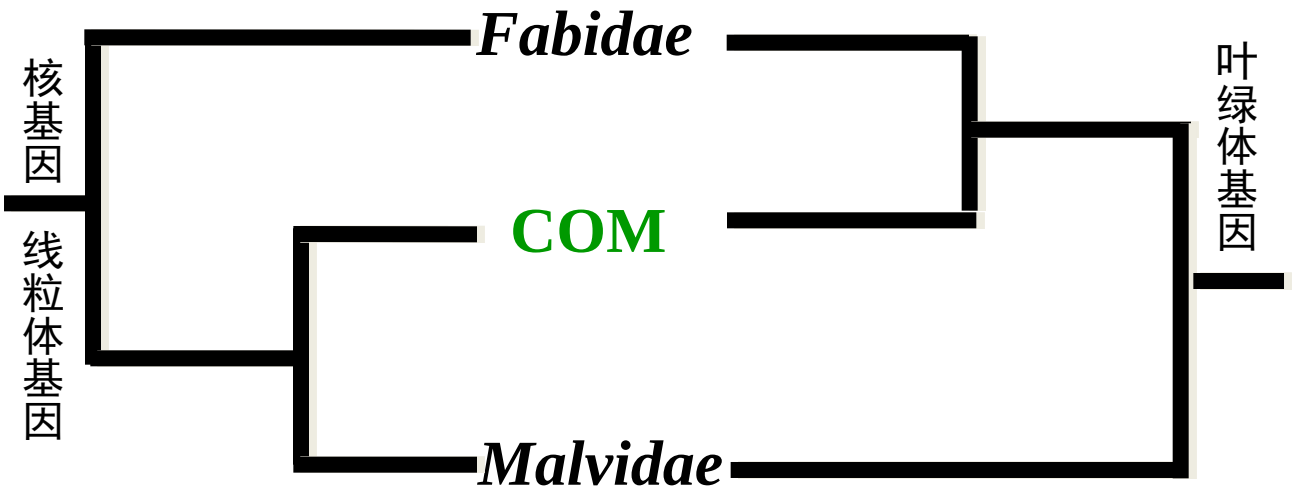
78 叶绿体基因单基因分析

COM支系统位置	基因数	详细
COM + *Fabidae*	36	BS > 60% 5个
COM + *Malvidae*	5	BS > 50% 0个
Fabidae + *Malvidae*	5	BS > 50% 1个81%
无信息	32	/

- 绝大部分叶绿体基因支持COM + *Fabidae*
- 线粒体基因*matR*和*rps3*支持COM + *Malvidae*
- 核基因都支持COM + *Malvidae*

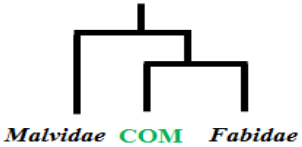
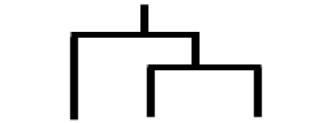
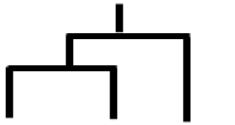
细胞器	基因	支持的拓扑结构	BS支持率
线粒体	*atp1*	未解决	/
	nad5	未解决	/
	matR	COM + *Malvidae*	34%
	rps3	COM + *Malvidae*	51%
核	*SMC2*	COM + *Malvidae*	24%
	MCM5	COM + *Malvidae*	30%
	MSH1	COM + *Malvidae*	50%
	MLH1	COM + *Malvidae*	50%
	SMC1	COM + *Malvidae*	80%

非软冲突

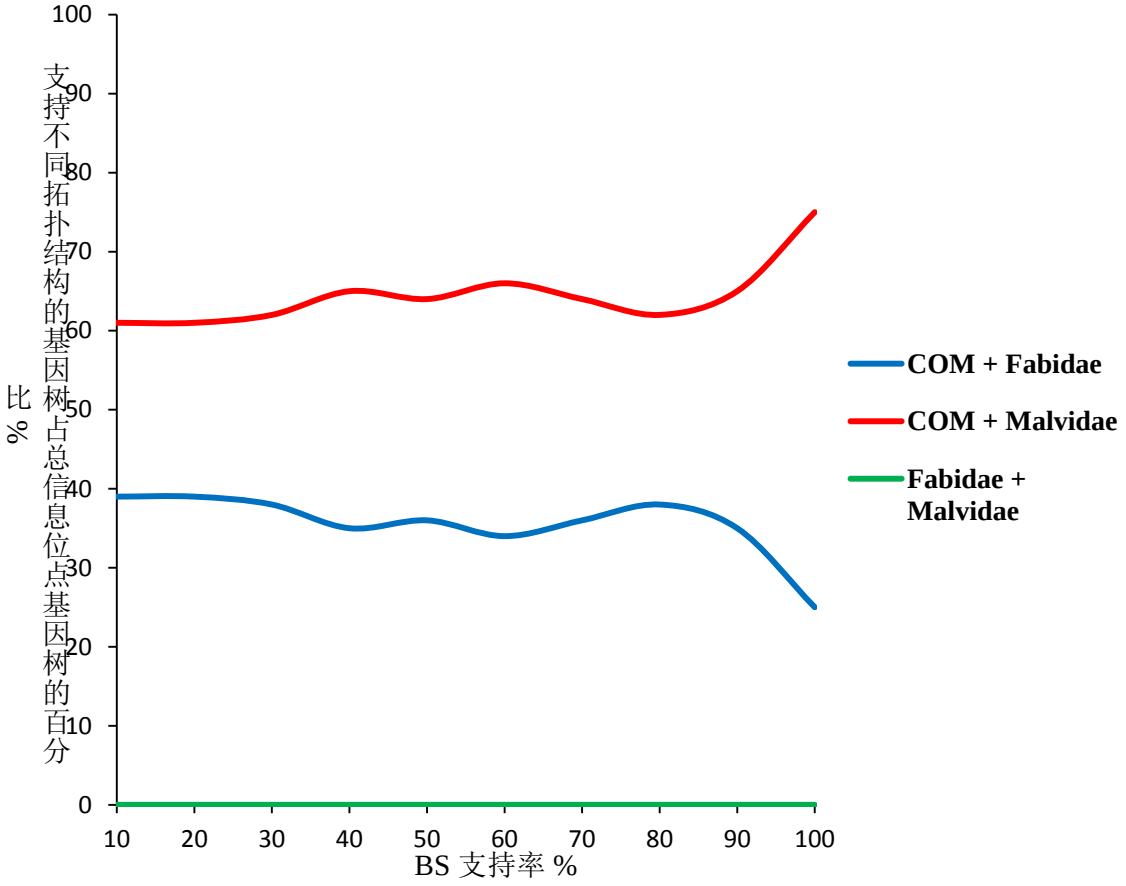


- 取样等同，分析方法一致
- 在目前的各种性状编码和“噪音”排除的测试方法中，COM支的系统发育关系没有受到系统误差和取样偏差的影响

单拷贝核基因分析结果

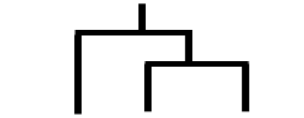
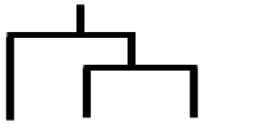
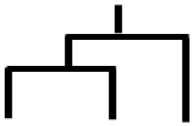
% BS支持率	 COM + *Fabidae*	 COM + *Malvidae*	 *Fabidae* + *Malvidae*
10	746 (39%)	1178 (61%)	0
20	449 (39%)	704 (61%)	0
30	283 (38%)	471 (62%)	0
40	171 (35%)	321 (65%)	0
50	115 (36%)	208 (64%)	0
60	68 (34%)	131 (66%)	0
70	49 (36%)	89 (64%)	0
80	29 (38%)	48 (62%)	0
90	15 (35%)	28 (65%)	0
100	3 (25%)	9 (75%)	0

单拷贝核基因分析结果

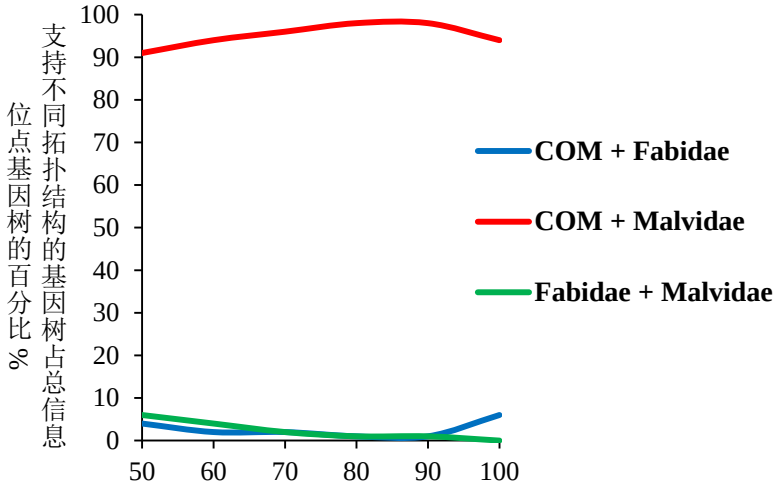


绝大部分单拷贝核基因支持COM + *Malvidae*

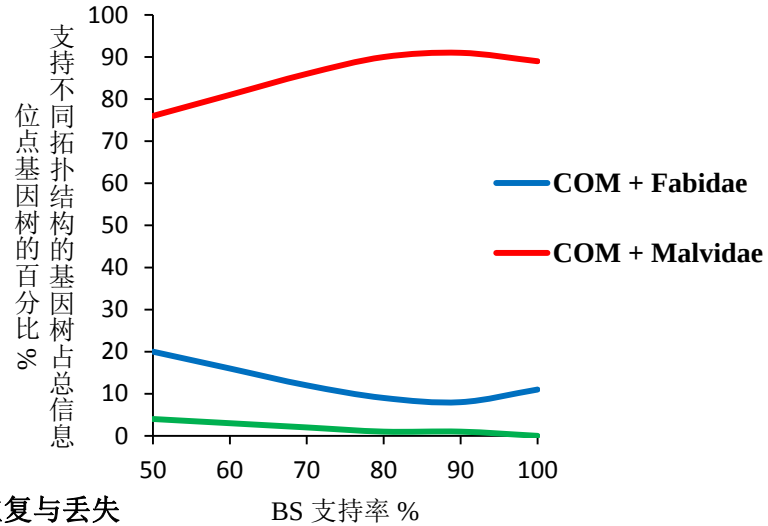
多拷贝核基因分析结果

进化事件	% BS 支持率	 COM + *Fabidae*	 COM + *Malvidae*	 *Fabidae* + *Malvidae*
基因重复	50	38 (3%)	973 (91%)	62 (6%)
	60	17 (2%)	723 (94%)	29 (4%)
	70	12 (2%)	515 (96%)	11 (2%)
	80	4 (1%)	308 (98%)	3 (1%)
	90	2 (1%)	155 (98%)	1 (1%)
	100	1 (6%)	15 (94%)	0 (0%)
基因重复与丢失	50	446 (20%)	1718 (76%)	82 (4%)
	60	267 (16%)	1390 (81%)	47 (3%)
	70	149 (12%)	1049 (86%)	26 (2%)
	80	72 (9%)	715 (90%)	11 (1%)
	90	30 (8%)	371 (91%)	5 (1%)
	100	7 (11%)	54 (89%)	0 (0%)
不完全谱系筛选	50	547 (27%)	1468 (71%)	40 (2%)
	60	358 (23%)	1160 (75%)	25 (2%)
	70	229 (20%)	884 (79%)	13 (1%)
	80	114 (16%)	598 (83%)	8 (1%)
	90	58 (16%)	299 (83%)	4 (1%)
	100	7 (13%)	44 (85%)	1 (2%)

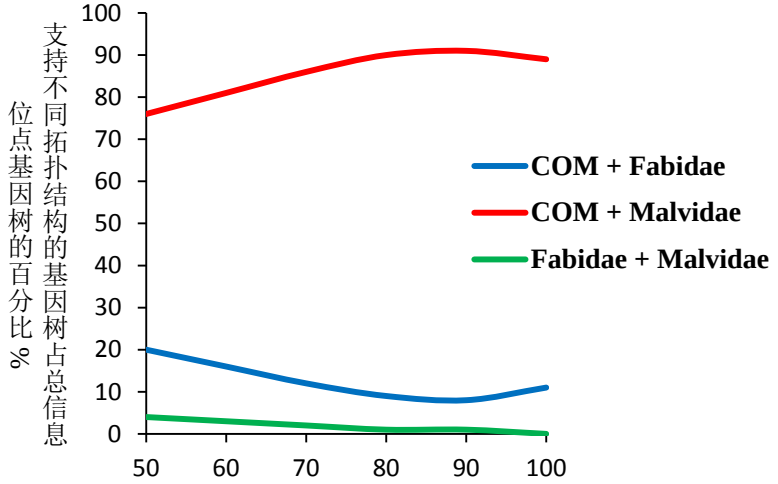
多拷贝核基因分析结果



基因重复



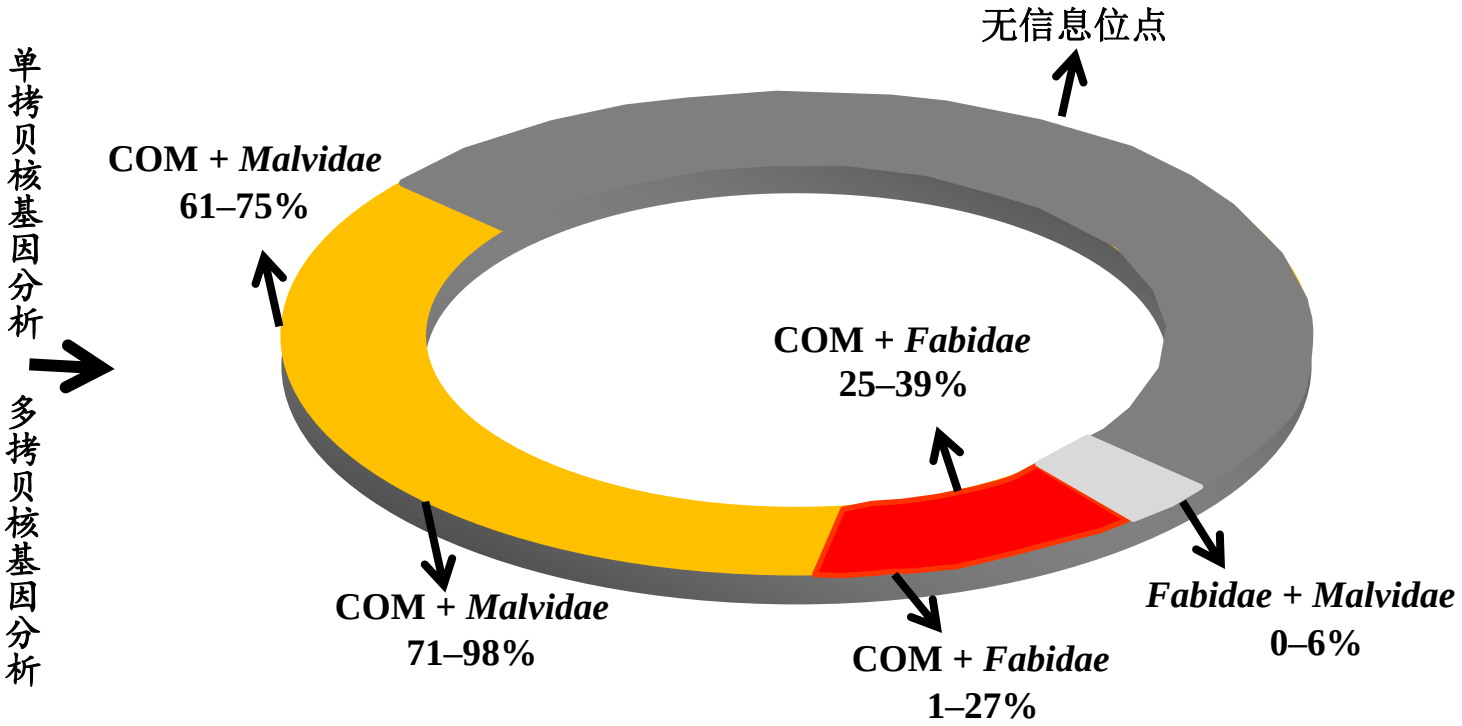
基因重复与丢失



不完全谱系筛选

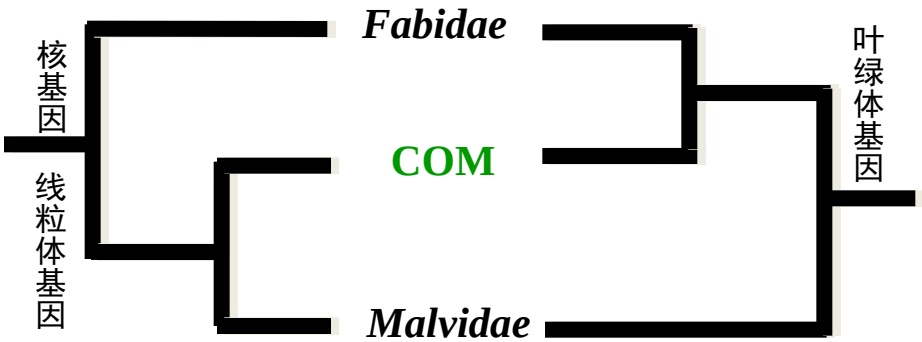
- >71%的多拷贝核基因支持COM + *Malvidae*
- *Fabidae* + *Malvidae*几乎无支持
- 三种进化事件都支持COM + *Malvidae*
- 基因重复事件对COM + *Malvidae*支持最高
- 不完全谱系筛选支持COM + *Fabidae*

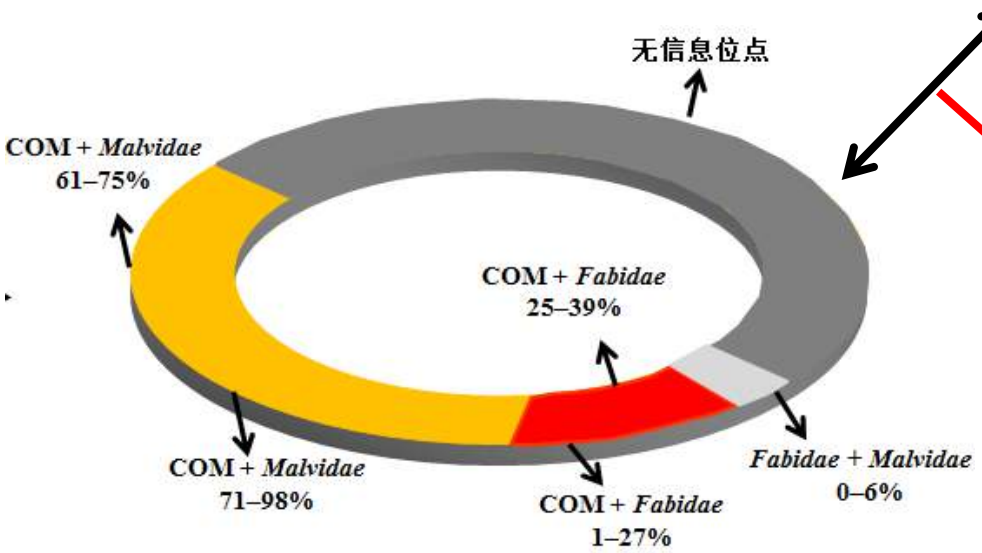
多拷贝核基因分析结果

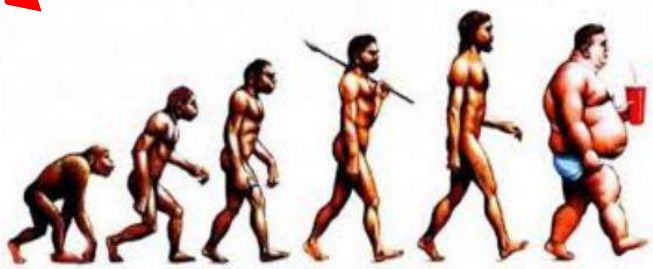


表明支持两大主要拓扑结构（COM + *Malvidae*和COM + *Fabidae*）基因的相对比例变化不大；COM + *Fabidae* 的系统发育信号不能忽略

冲突原因


核基因
线粒体基因
Fabidae
COM
Malvidae
叶绿体基因


无信息位点
COM + Malvidae
61–75%
COM + Fabidae
25–39%
COM + Malvidae
71–98%
COM + Fabidae
1–27%
Fabidae + Malvidae
0–6%



生物过程

不完全谱系筛选

古杂交

假说

不完全谱系筛选

- **蔷薇类两大支的分化时间**

  Fabidae 108–91 Mya

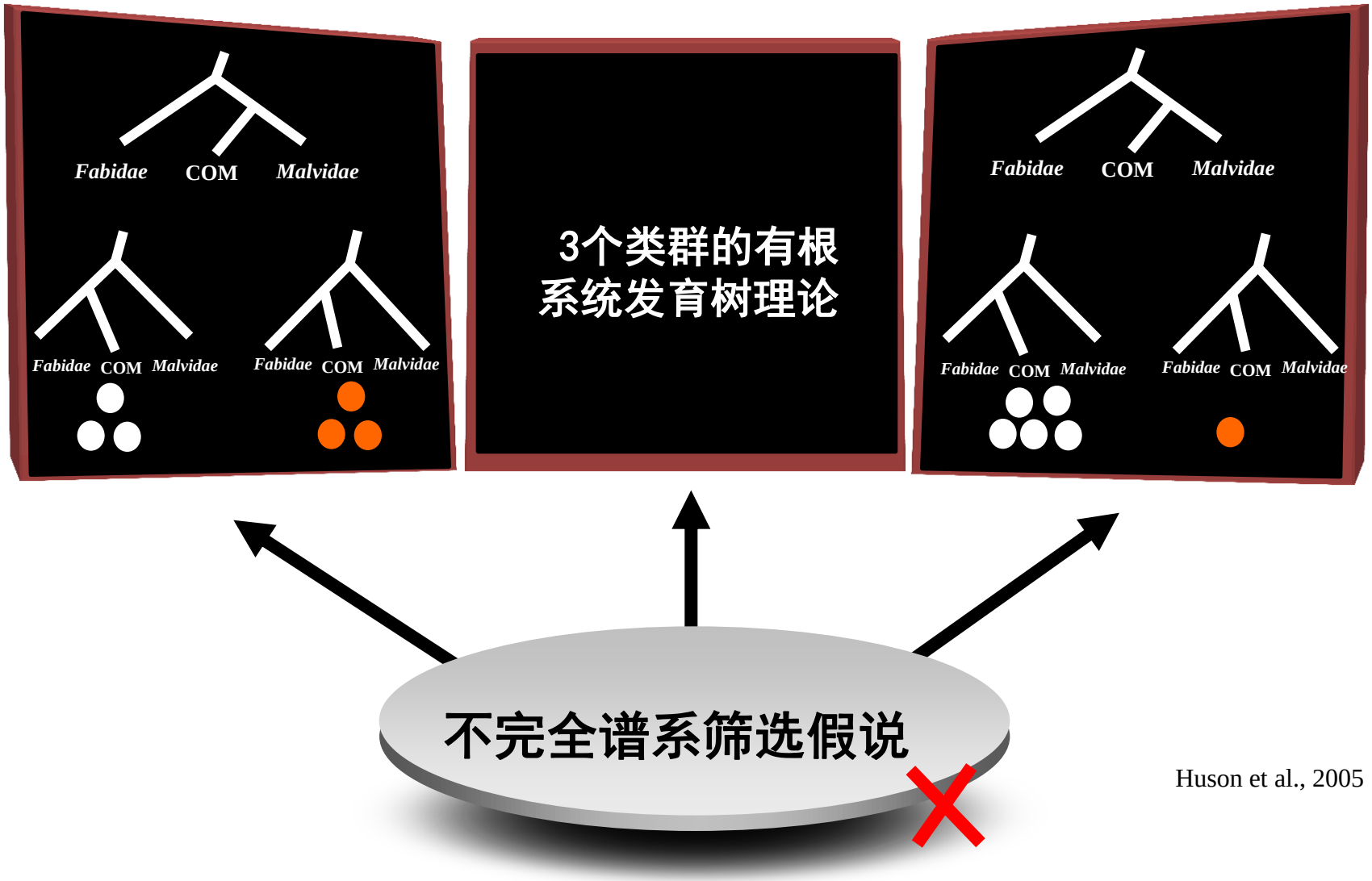
  Malvidae 107–83 Mya

- **不到4–5 Mya 短暂的时间内，蔷薇类迅速完成了物种分化**

- **多发生在物种和居群水平**

Wang et al., 2009

不完全谱系筛选



Huson et al., 2005

不完全谱系筛选

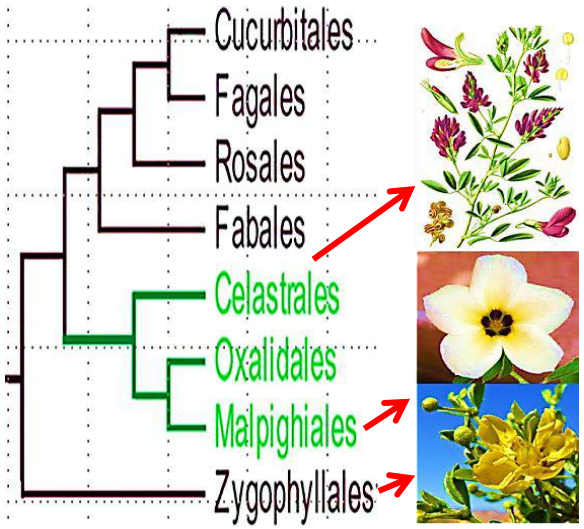
- 很难区分不完全谱系筛选和杂交
- 久远的蔷薇类分化时间
- “三个类群有根的系统发育关系”理论
- 核基因组数据的进一步丰富

Rosenberg & Nordberg, 2002
Degnan & Rosenberg, 2006

古杂交

古杂交

1. 植物类群繁衍中多次经历过杂交、渐渗；各阶元很普遍 20年前，报道100多例。
2. 被子植物的遗传模式多样
3. 蔷薇类内叶绿体父系遗传
4. 实验证据



紫苜蓿

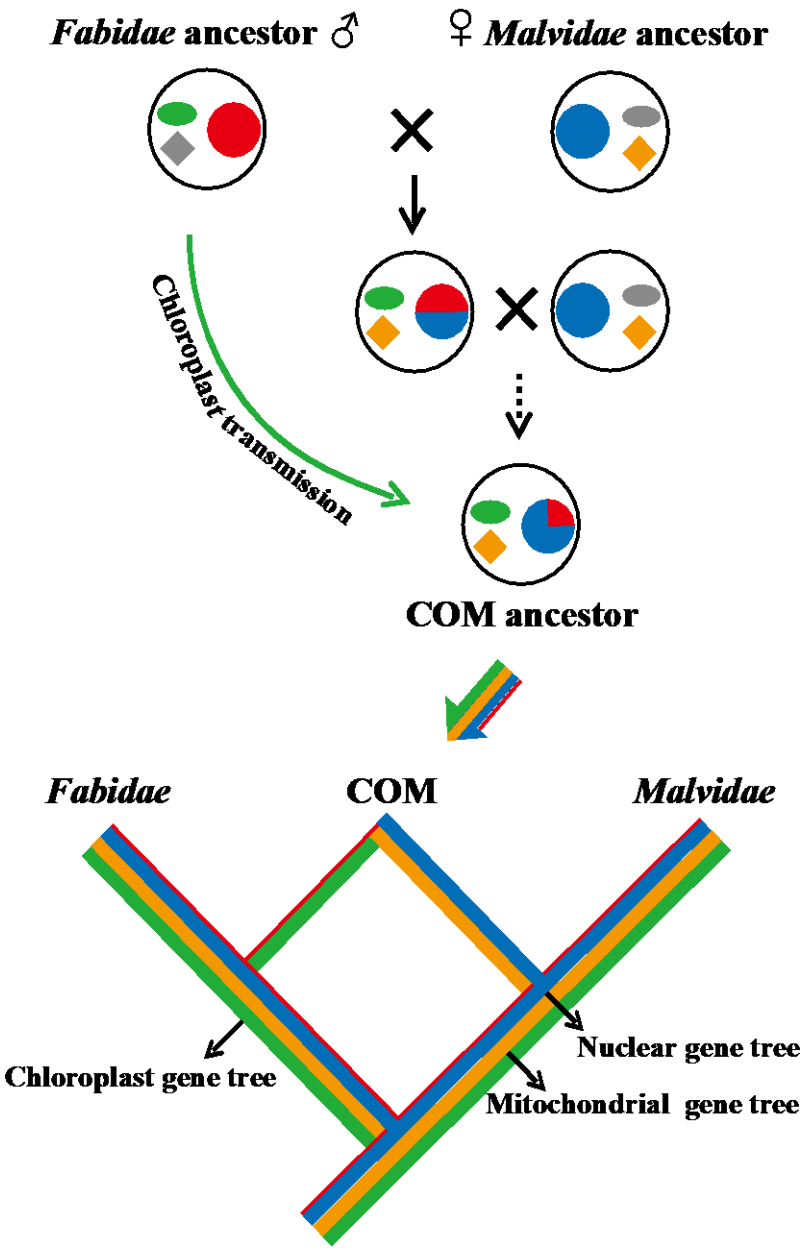
时钟花

石碳酸灌木属

古杂交

Corriveau & Coleman, 1988
Masoud et al., 1990
Rieseberg et al., 1996
Shore & Triassi, 1998
Yang et al., 2000

古杂交


Fabidae ancestor ♂
♀ Malvidae ancestor
Chloroplast transmission
COM ancestor
Fabidae
COM
Malvidae
Chloroplast gene tree
Nuclear gene tree
Mitochondrial gene tree

报告内容

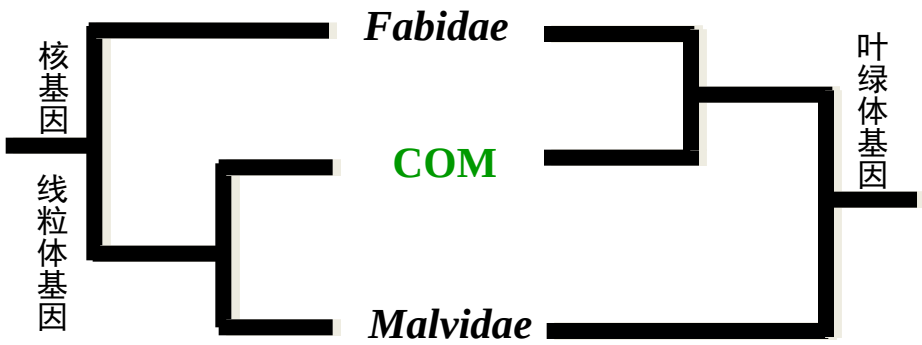
1 研究背景

2 研究方案

3 结果与讨论

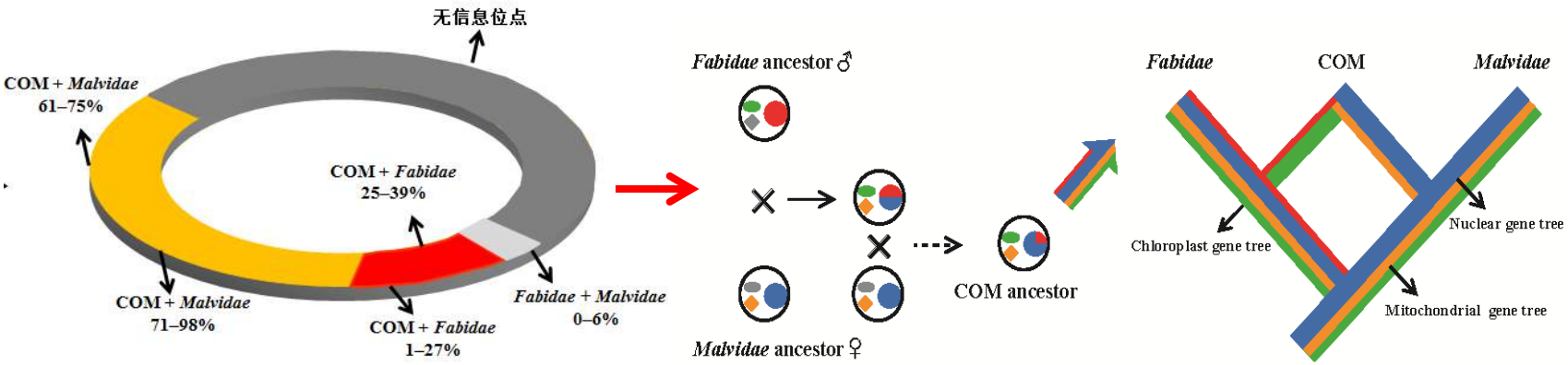
4 结论与展望

结论



对三个取样近等同的叶绿体、线粒体矩阵进行了多种性状编码和“噪音”排除的系统发育关系验证分析，重现了COM支冲突的系统发育关系；这种冲突的系统发育关系并没有受到取样偏差或者系统误差的影响，可能是生物过程造成的

结论



对单拷贝和多拷贝核基因组数据分析发现，核基因不等比例地支持COM支不同的系统发育关系，进一步揭示了早期的蔷薇类在快速辐射分化的过程中，*Fabidae*和*Malvidae*祖先可能发生了古杂交并随叶绿体基因组渐渗的进化事件。由此造成了关于COM支的叶绿体与核、线粒体基因树间的冲突，最终形成了核基因组内两种相互冲突且代表亲本遗传信息的系统发育信号共存的格局

展望

- 基因组数据和系统发育基因组学方法在检测、识别并解决生物进化事件导致生命之树深层次系统发育关系的冲突方面是行之有效的

- 随着海量数据与日俱增。如何应对这些指数增长的数据，选择合适的基因信息，掌控数据处理的精度，可靠而合理回答地科学问题，揭示生命进化的“恼人之谜”是人们当前面临的首要任务

致 谢！

本论文得到国家自然科学基金委面上项目基金（NNSF 31270268）、国家重点基础研究发展计划（No. 2014CB954101）及中国科学院外国专家特聘研究员计划（No. 2011T1S24）的联合资助，特此致谢！

感谢导师陈之端研究员的指导和培养，及组里成员的包容和支持！

感谢弗罗里达大学Douglas Soltis、Pamela Soltis 和Gordon Burleigh在学术上的指导和熏陶！

感谢系统中心各位老师和同学在我博士期间给予的帮助和关爱！

再次感谢各位评委老师和同学们参加我的博士毕业论文答辩！


欢迎各位老师同学批评、指正！