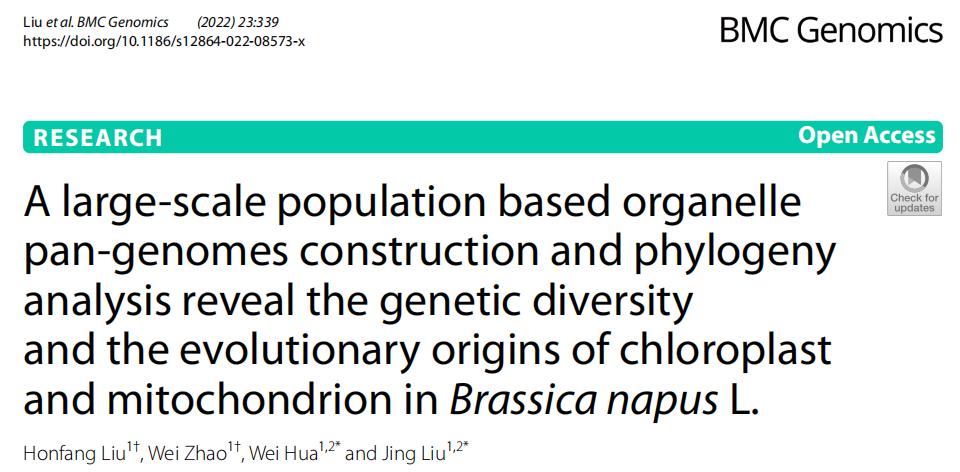
## 标题

泛基因组案例 |基于大规模群体的细胞器泛基因组构建和系统发育分析揭示了油菜叶绿体和线粒体的遗传多样性和进化起源



## 导读

异源四倍体油菜是世界上重要的油料作物.由于缺乏野生资源,油菜籽的来源仍未确定.尽管已对很多芸薹属核基因组进行了一定的遗传结构和系统遗传学研究,但全局模式下的细胞器基因组信息在很大程度上仍是未知的,而细胞器基因组可以为油菜系统发育研究提供独特的材料.在此,基于全球收集的1579份油菜材料的从头组装,作者构建了油菜叶绿体和线粒体泛基因组,并研究了油菜与青菜,白花甘蓝间的遗传多样性、系统发育关系.

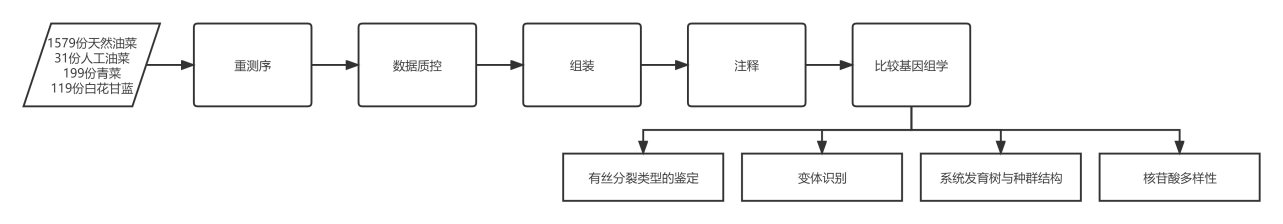
## 论文ID

原名：A large-scale population based organelle pan-genomes construction and phylogeny analysis reveal the genetic diversity and the evolutionary origins of chloroplast and mitochondrion in Brassica napus L.  
译名：基于大规模群体的细胞器泛基因组构建和系统发育分析揭示了油菜(Brassica napus L.)叶绿体和线粒体的遗传多样性和进化起源  
期刊：BMC genomics  
IF：3.969  
发表时间：2022.4  
通讯作者：华玮/刘静  
通讯作者单位：中国农业科学院油料作物研究所/农业农村部油料作物生物学与遗传育种重点实验室,湖北武汉,430062  
DOI：10.1186/s12864-022-08573-x

## 实验设计

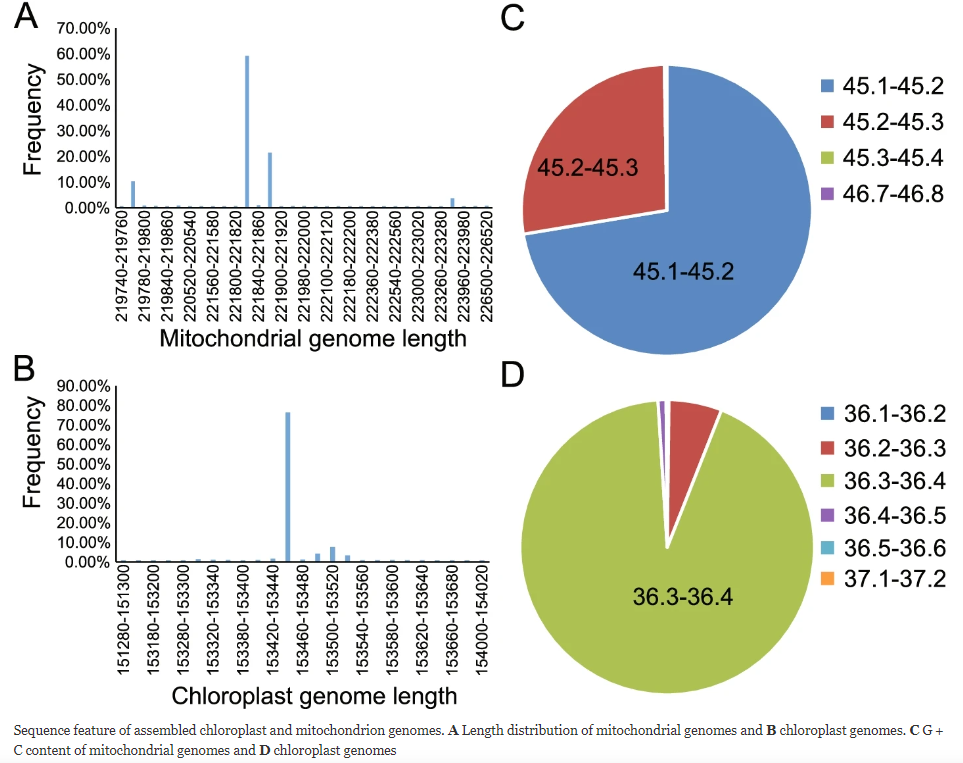
研究材料：1579份天然油菜、31份人工油菜、199份青菜、119份白花甘蓝  
技术方法：NCBI下载测序数据/CTAB 法从叶片提取DNA二代建库测序

## 实验路线

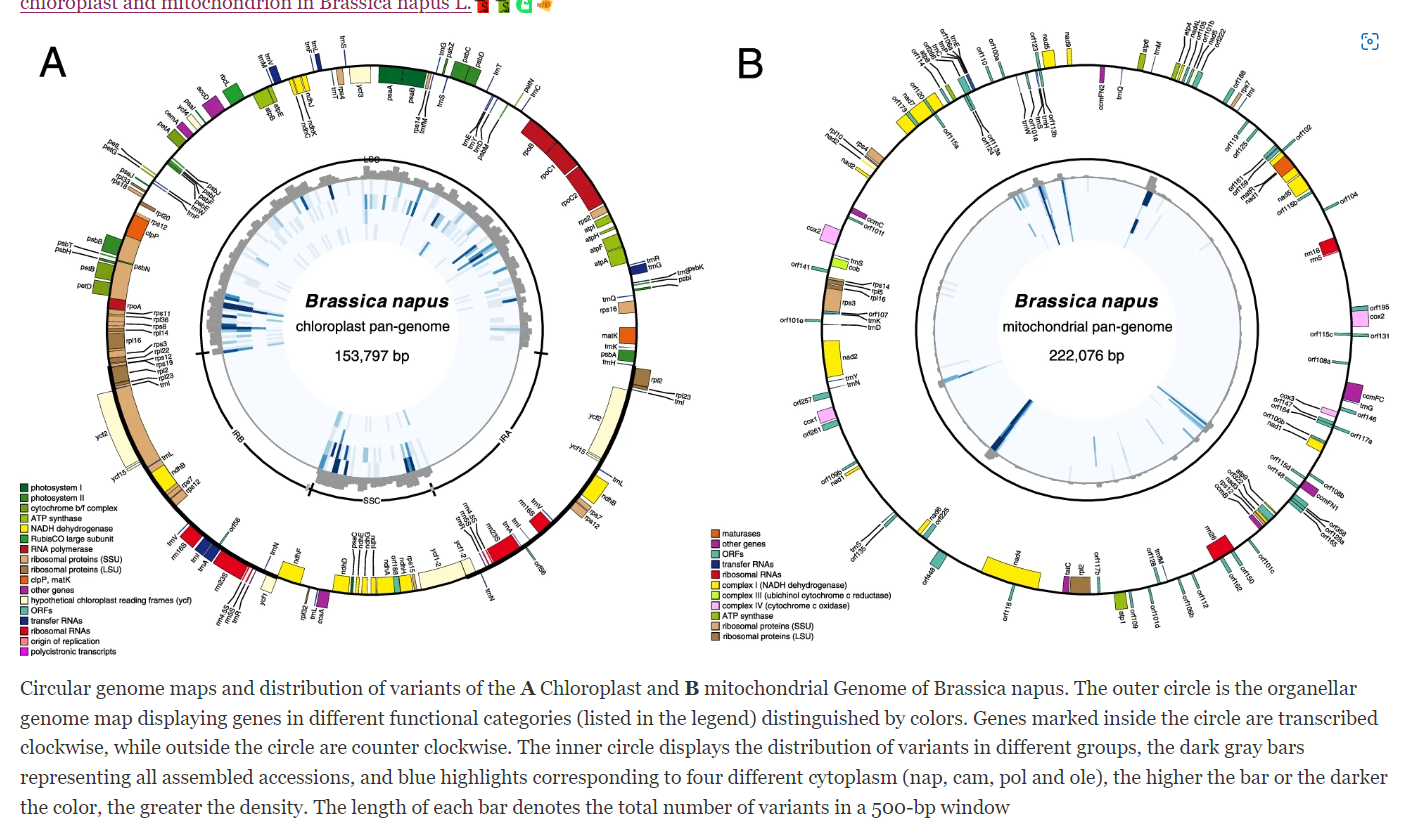


## 结果

### 细胞器基因组的组装和注释



1579份油菜籽样品的基因组测序数据来自NCBI数据库SRP155312和PRJNA358784.在对reads两端的低质量区域进行质量检查和修剪后,先将clean data分别映射到35个mt基因组组成的线粒体基因组序列簇和42个cp基因组组成的叶绿体基因组序列簇.映射的双端reads分别用NOVOPlasty ARC软件进行提取和从头组装成cp和mt基因组.样本的提取数据大小不同(mt为20至400 Mb,cp为11 Mb至1.8 Gb),但均达到较高水平,平均覆盖率(> 100 ×).  
细胞器基因组(cpDna和mtDNA)每个单独组装.总计1327个cpDna和1456个mtDNA被组装成环状分子,并在下游分析中予以考虑.cp基因组相对集中的大小为153 kb,而mt基因组的大小在219到226 kb之间(图1A,B).相应地,cp和mt基因组的主要GC含量分别约为36.3%和45.2%(图1C,D).在所有组装的序列中,47个cpDna和225个mtDNA含有gap(以N填充),其长度在1至770 bp之间.

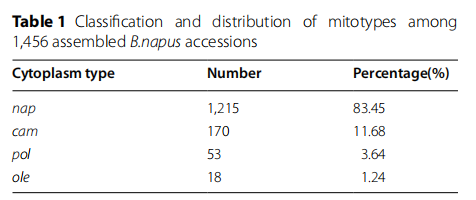


用GeSeq对组装的基因组进行了注释.使用由从已知芸苔属物种获得的基因组作为参考.

cp泛基因组是一个153,797 bp的环状分子,由两个反向重复区以及一大一小单拷贝(LSC和SSC)区域组成.共鉴定出87个蛋白编码基因,3个功能未知的orf,8个rRNA和37个tRNA序列(图2A),约占基因组的49%.

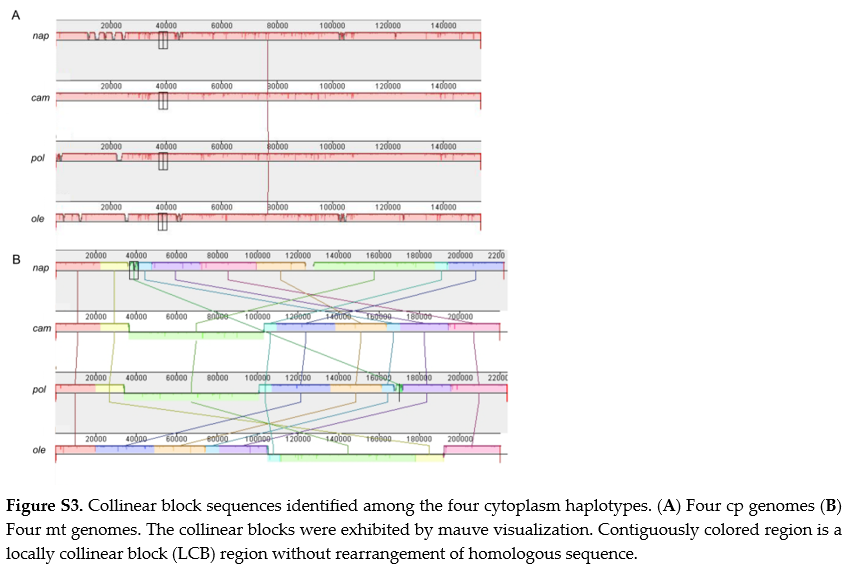
mt泛基因组的总大小为231,901 bp,由一个222,076 bp的环状分子和5个平均长度为1,965 bp的可分配scaffolds组成(图2B).共有35个蛋白编码基因,注释出3个rRNA、18个tRNA和62个orf,占基因组序列的28%.其中15个orf具有有丝分裂特异性,包括CMS相关基因orf224.

### 细胞质单倍型的鉴定



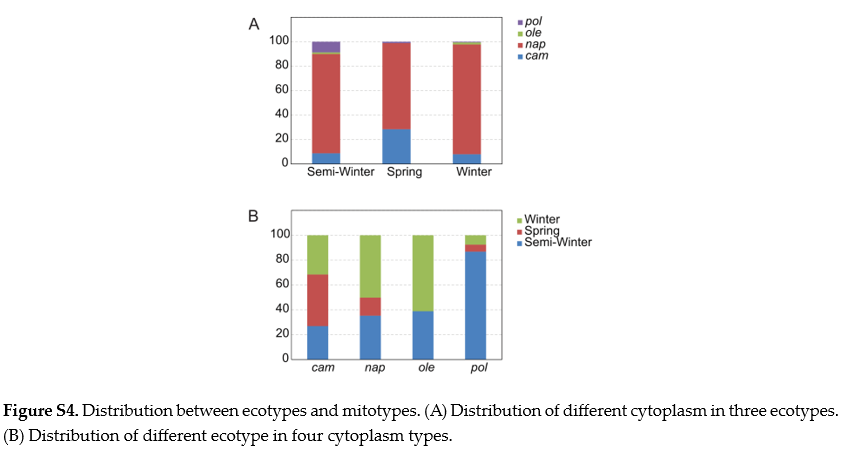
为了区分所有材料的细胞质类型,先前的研究者开发了12个有丝分裂特异性标记(基因组特异性序列,>100bp)和34个同源ORFs,而这些orf在六种芸苔属有丝分裂类型的mt基因组中是不一致的.研究人员为样本定义了完全符合线粒体类型与MSSs/orf之间关系的细胞质类型.基于结构比较,符合混合MSS和orf相应条件的样本被定义为“相似”类型.

在1456份材料中检测到四种主要的有丝分裂类型.nap细胞质1215例,cam细胞质170例,pol细胞质53例,其余18例为ole细胞质(表1).nap型材料占总数的83%,表明nap有丝分裂型在天然油菜中占优势.



细胞质类型相同的材料表现出由Mugsy比对结果支持的一致的细胞器基因组结构.为了检查细胞器基因组的结构特征,使用Mauve软件对来自不同细胞质的四个cp和四个mt基因组进行多基因组比对和大规模重排鉴定(图S3).

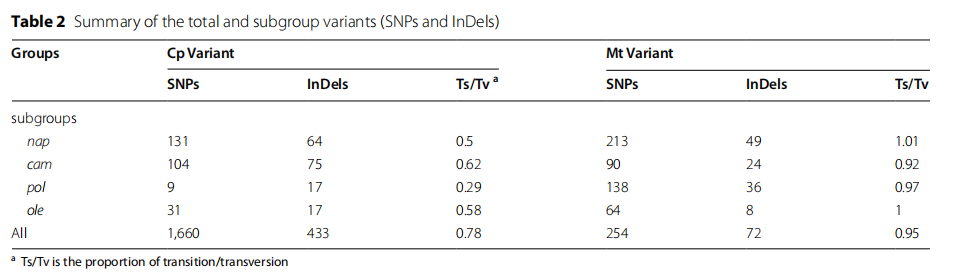
来自不同细胞质的cp基因组在整个基因组上表现出共线性.然而,这四种有丝分裂类型的mt基因组却发生了重排.图中可以发现cam和pol完全共线,因为它们以一致的顺序和方向共享所有相同的共线性区块,但在任何其他两个mt基因组之间至少发生了三次重组事件.



为了探索生态型和有丝分裂型之间是否存在一定的相关性或偏差,研究人员分别建立了不同有丝分裂型在三种生态型中的分布和不同生态型在四种有丝分裂型中的分布(图S4).3个生态型间有丝分裂型分布无明显差异,与总体分布相似,除了cam型占春季组中28%,在其他3组则少于10%.

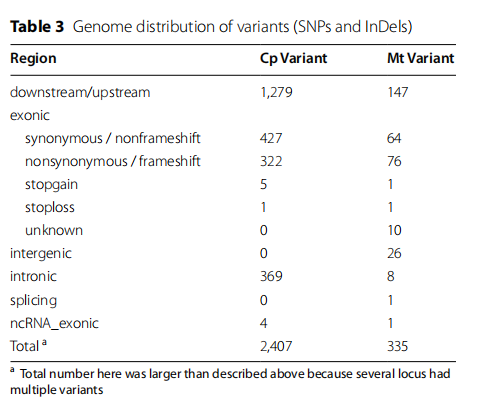
然而,三种生态型在不同有丝分裂类型中的比例不一致.pol型材料以半冬型为主(83.63%),ole型材料仅检出两种生态型(冬型占61.11%,半冬型占35%),cam中三种生态型的数量相近,nap材料中的半冬、春、冬生态型比例为2:1:4.

### 芸苔属植物的全基因组变异



在组装的cp和mt基因组中分别检测到4115和450个原始变异体.去除单个样本特异性变异体后,从cp基因组中获得了2092个高质量变异体,包括1660个SNPs(79.35%)和433个InDels(20.65%).此外,在mt基因组中鉴定出254个SNPs (77.91%)和72个indel(22.09%),相应地在cp和mt基因组中达到每kb13个和1.5个变异体的密度(表2).

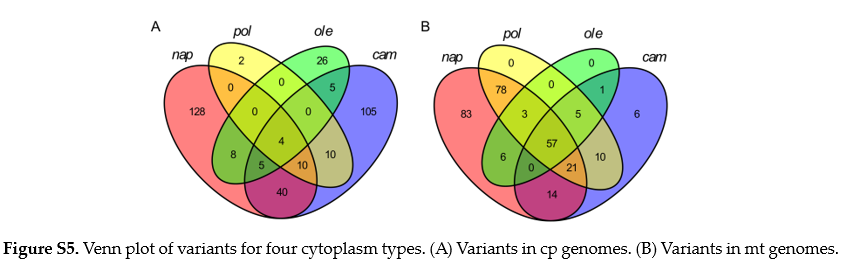
在整个群体中,cpDna中检测到的SNP和indel数量远远高于mtDNA中检测到的数量,但在cpDna中其每个亚组中的变异体数量较低,表明不同有丝分裂型的SNP/小indel在mt中比在cp中更丰富.在cpDna中未检测到结构变异.



cp变体沿参考基因组均匀分布,除了两个反向重复区,由于多次比对,其reads被跳过(图2).在总共1327个cp变体中,53%位于上游/下游区域,11%位于内含子,31%的变体位于编码区域,其中43%被预测为非同义或移码突变(表3),可能导致不同的功能蛋白编码.变体沿基因组的分布与四种细胞质类型的总体分布一致.

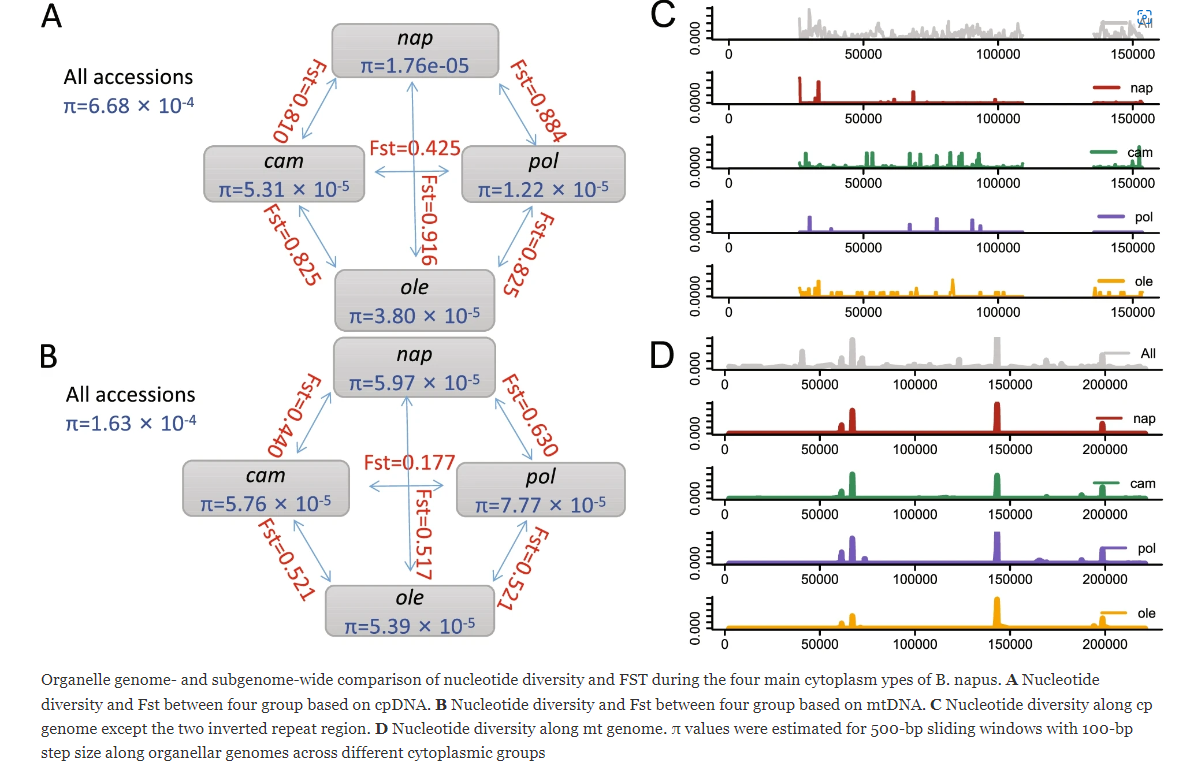
mt变体分散在整个参考基因组中,但在编码区富集(P值 < 0.0001,超几何分布phyper检验).值得注意的是,45%的变体位于编码区,这部分区域在所有编码区中,其构成的频率远高于基因组序列的28%.在编码区的变体中,58%被预测为受功能影响.

在四种细胞质类型中检测到的cp变异体的比例与mt不同.此外,还观察到转换/颠换(Ts/Tv)比率不一致(表2).在mt基因组中四种有丝分裂类型的比例约为1,而在cp基因组中的比例在0.29至0.78之间,表明Ts/Tv在cpDna中的比例低于mtDNA.



四个亚组(nap、pol、cam和ole有丝分裂型)中有4个常见cpSNPs和57个常见mtSNPs.除pol外,每种基因型的大部分cp变体是特异性的,而只有一小部分mt变体是特异性的(图S5).

### 细胞质基因组的多样性分析



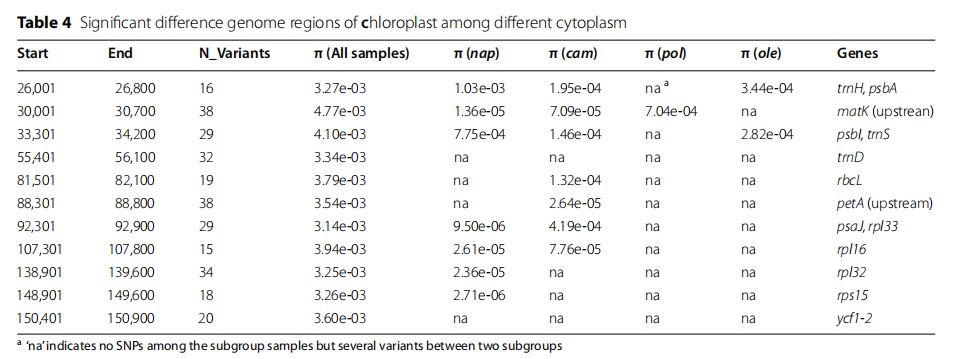
基于全基因组变体对cp和mt基因组进行了核苷酸多样性(π)和固定指数(Fst)分析.

研究人员发现cp和mt基因组中的每个亚组具有相似且较小的平均核苷酸多样性.最高的多样性是mt基因组的pol中的丰度为7.77×(10^-5),cp基因组nap中最低为1.76×(10^-5)(图3A,B).但在所有材料中差异较大,因为完整组装的cpDna和mtDNA的核苷酸多样性分别为6.68×(10^-4)和1.63×(10^-4).

另一方面,除了cam和pol组之外cp基因组中每两个群体之间的遗传距离(Fst)在0.810至0.916之间,mt基因组中在0.440至0.630之间(图3A、B),显示这4个细胞质组之间高度分化,这在PCA图中显示得更直观(图4A,D).在mt和cp中,cam和pol之间的分化程度均低于其他组之间的分化程度,表明存在密切的进化起源.

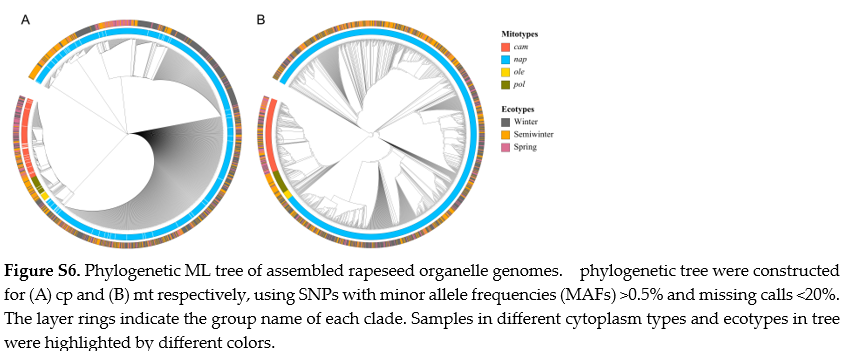
cp基因组表现出的组内核苷酸多样性小于mt基因组,但具有较大的组间差异,这意味着不同细胞质类型之间的cp基因组差异体现为单碱基多态性,同时保持了与线粒体基因组不同的保守基因组结构.

基于500bp滑动窗口和100bp步长估计了四种细胞质类型(nap、cam、pol和ole)中细胞器基因组的多样性.mt基因组各亚组间结果一致(t检验P值> 0.05),而cp基因组间存在显著差异(t检验P值< 0.01)(图3C,D).



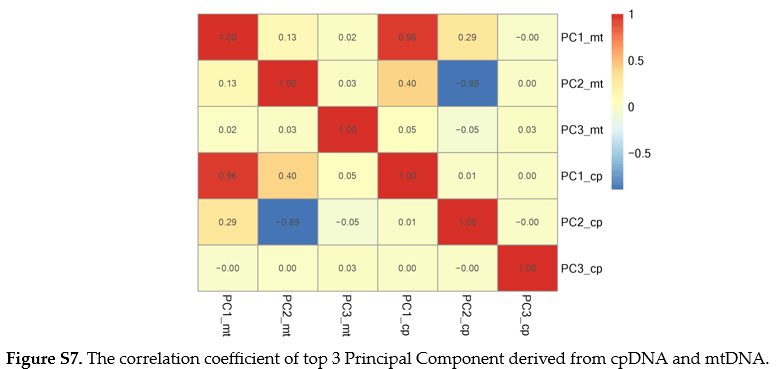
四个亚组中的两个亚组之间有几个区域的多态性较高.为了检测这些基因组区域,研究人员计算了多样性减少(ROD)值,基于整个组与nap、cam、pol和ole各组间的核苷酸多样性比值,使用500bp窗口.具有最高1% ROD值的区域被排除,连续窗口被合并成11个区域(表4).

### 遗传结构和系统发育关系

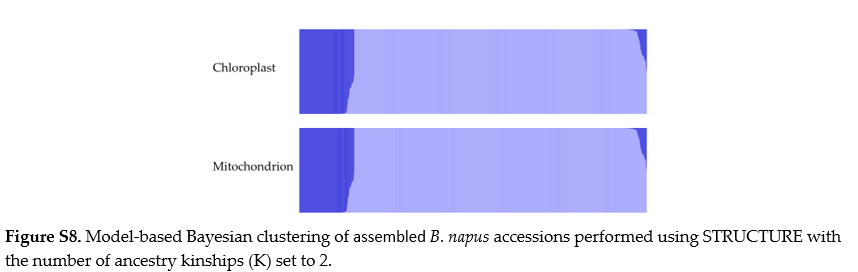


为了研究油菜及其二倍体亲本与人工的油菜籽细胞器基因组之间的遗传结构和系统发育关系,通过来自油菜细胞器泛基因组的定位reads检测199份青菜和119份白花甘蓝材料,以及31份合成油菜材料(11份来自未发表的测序数据)的SNPs.接下来,在筛选出油菜中等位基因频率较小(MAF) < 0.02,同时满足在保留物种特异性位点的三个物种中MAF < 0.005的位点后,研究人员进行了主成分分析(PCA)以及系统发育树和种群结构分析.

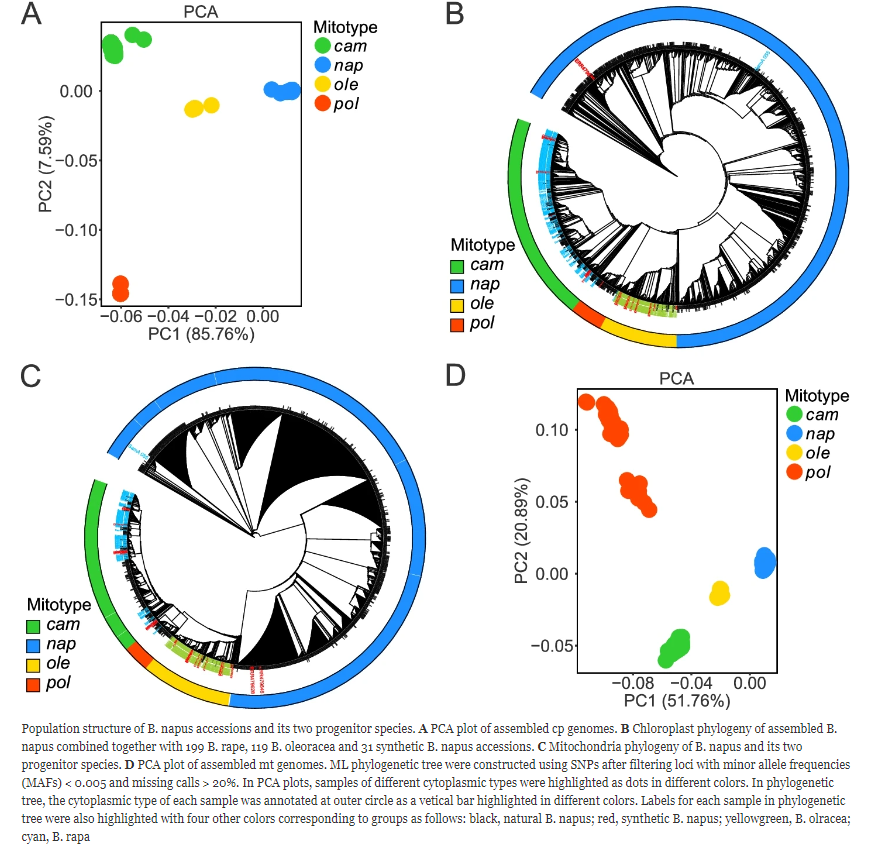
基于PCA和最大似然(ML)树结果,研究人员发现具有相同细胞质类型细胞器基因组的芸苔属材料聚集在一起,在cpDna衍生树和mtDNA衍生树中具有几乎相同的进化位置,但与主要受气候区和纬度差异影响导致的生态分型无关(图S6).



来源于cpDna和mtDNA的相关系数PC1为0.96(图S7),表明油菜自然群体中的cp和mt基因组同时进化,PC1方差比例对于cpDna为85.76%,对于mtDNA为51.76%.

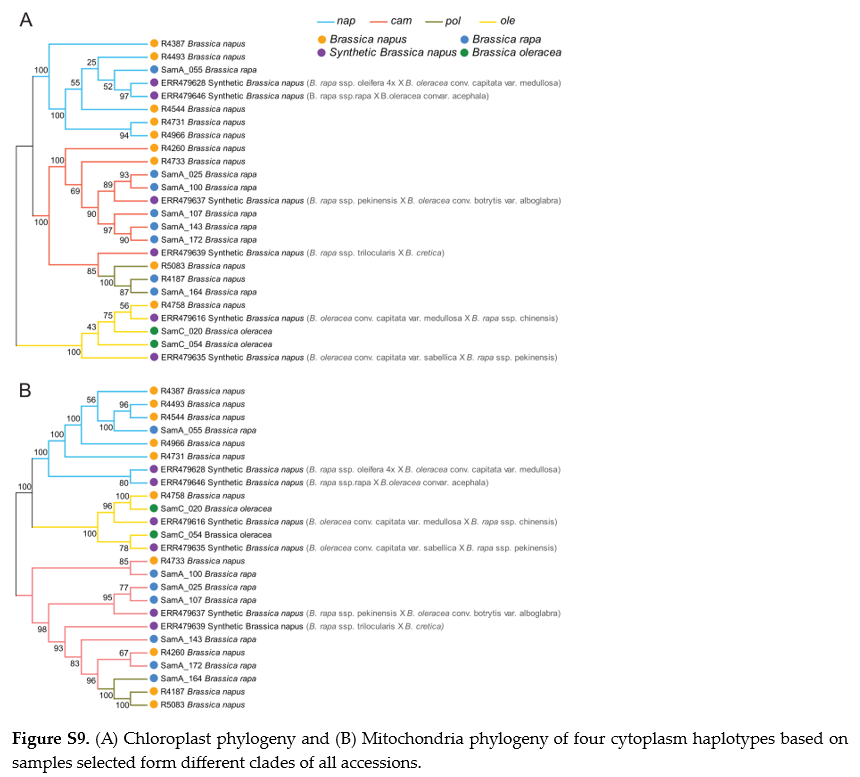


根据能区分nap和其他有丝分裂类型的群体结构结果,确定油菜材料的最佳群体聚类数为K = 2(图S8).



然而,在PCA图中,通过PC1和PC2,油菜进一步聚类为4组(图4A,D),与nap、cam、pol和ole有丝分裂类型完全一致.在cp PCA图中,nap和cam分组PC1水平接近.  
为了研究三种芸薹属种内和种间不同细胞质类型的驯化历史,研究人员构建了一个包含青菜、白花甘蓝、合成和天然油菜群体系谱的树.mt和cp树显示了相似的种群结构.油菜材料分化为3个分支,揭示了油菜的复杂起源(图4B,C).在先前的研究中,少数油菜材料与大多数青菜归为一类.在这里,几乎所有的青菜都被归为最接近油菜cam的一组,其次靠近油菜pol,揭示cal/pom油菜可能是由青菜的祖先进化而来.此外,所有的白花甘蓝材料都聚集在ole一分支中,表明另一个独立的细胞器来源于白花甘蓝(图4B,C).

在自然油菜中占优势的nap有丝分裂材料,没有与青菜或白花甘蓝聚类在一起.但是,1个青菜(SamA\_055, ssp. pekinensis)和2个合成的油菜材料(ERR479628 /ERR479646),它们的母本来自青菜.油茶4 ×和青菜B. rapa ssp. rapa被归到nap,而其他母本为青菜的人工材料被聚类到cam分支中,而其母本来自白花甘蓝的材料被聚类到ole分支中.同时,基于MSS图谱将这3份材料检测为nap细胞质,支持了聚类评估结果.



上述信息表明,nap细胞质可能是从青菜的细胞质进化而来,这在过去的研究中被认为是有争议和不明确的.为了更清楚地研究不同类型细胞质之间的系统发育关系,研究人员从每个分支中选择了少量代表性材料,并构建了系统发育树(图S9).

## 结论

细胞器基因组对细胞质遗传相关农艺性状的形成具有重要作用,是研究油菜母源的有效手段.

本研究以全球收集的油菜材料为基础,构建了油菜叶绿体和线粒体全基因组,并鉴定了全基因组的方差和多样性,为油菜细胞质遗传相关性状的研究提供了材料,为理解油菜细胞器基因组的整体多态性提供了信息.

包含青菜,白花甘蓝以及天然/合成种群油菜的系统发育树揭示了不同油菜细胞质单倍型的起源.在青菜群体和以青菜为母本的合成油菜中鉴定出nap型这一情况提供一个新观点,表明青菜可能是nap型油菜的母体祖先.

细胞质单倍型鉴定和系统发育为油菜细胞器基因组进化提供了新的视角.

## 小编点评

在很多文章里,研究人员会看到研究者在研究一种植物起源进化的时候,通常会选择叶绿体基因组来入手,因为叶绿体基因组相对保守,组装也较方便.

在这篇文章里,作者不仅使用了叶绿体基因组,还构建了叶绿体与线粒体的泛基因组,更重要的是其数据规模极为庞大,相较而言研究比较全面.除此之外,作者还进行了细胞质单倍型的鉴定,主成分分析,fst分析等,让结果从多方面来互相验证,比只用进化树来分析结果更好,更有说服力.

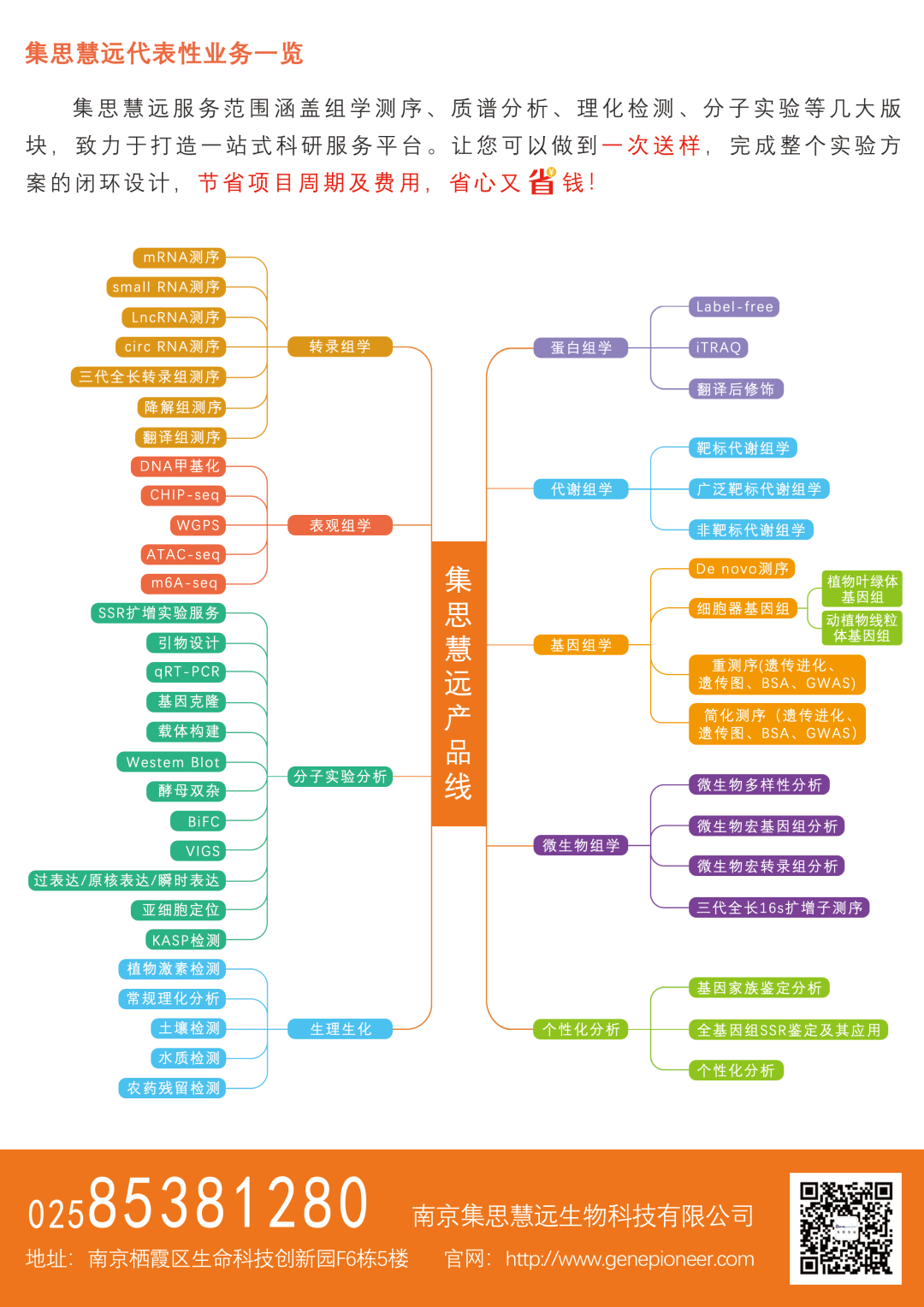
## 对标产品介绍

南京集思慧远生物科技公司在植物叶绿体线粒体以及动物线粒体基因组测序组装组注释中已积累丰富的经验,先后测序了几百种植物叶绿体线粒体基因组,同时公司还有线粒体基因组组装专利,是目前国内细胞器基因组测序的“领航”公司之一.公司目前已发表细胞器文章200余篇,分别发表在不同的杂志上,如《Genomics》、《Horticulture Research》、《International Journal of Molecular Sciences》、《Ecology and Evolution》、《Frontiers in Genetics》、《BMC Plant Biology》、《genes》等.公司的细胞器基因组测序组装周期短,组装准确、成本低.动植物细胞器基因组只需一篇叶子,一只昆虫又或者物种肌肉组织如肝脏、胸肌,提取下物种DNA ,可在1.5个月-3个月完成测序组装分析.如果你想在短期内发一篇2-3分的SCI 文章,那欢迎来电咨询：025-85381280.

## 文献链接

文章在线链接<https://doi.org/10.1186/s12864-022-08573-x>,如想获取pdf版本,可以联系小编微信

## 产品列表



## 公司简介

南京集思慧远（GENEPIONEER）生物科技有限公司于2015年4月成立,是一家专业从事高通量生物检测服务的企业.公司秉承“创新生物科技,服务惠及于民”的宗旨,依托前沿的基因测序技术和大数据分析平台,逐步建立一整套完善的分子生物学实验平台,打造真正的生命科学高新技术研发中心,加强科技创新和服务创新,推进生命科学、农业育种、食品安全和医疗健康等领域的研究与应用,致力于通过不懈的努力,成为国际生命科学领域的航空母舰,造福全世界,惠及全人类.  
公司目前业务涵盖各组学测序服务、生理生化检测服务、分子实验外包服务以及生物大数据分析服务.主要产品涉及基因组学、转录组学、蛋白组学、代谢组学、表观组学、微生物组学等六大组学领域.公司2017年获得国家高新技术企业称号.公司自主开发各类实验技术以及数据分析算法,已获批软著62件,授权发明专利1件,受理发明专利5件.已为国内500多家高校、农科院、医院、医学院、药企等各类研究机构提供过科研服务.