

研究开放存取



# 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遗传多样性以及甘蓝型油菜叶绿体和线粒体的进化起源

Honfang Liul, Wei Zhao1, Wei Hual, 2\*和 Jing Liul, 2\*

## 摘要

背景：异源四倍体油菜是世界重要的油料作物。由于缺乏野生资源，油菜籽的来源仍未确定。尽管已对大组甘蓝 B. 菜青虫和甘蓝。

结果：基于有丝分裂类型特异性标记和有丝分裂类型变异 orf，在我们的组中鉴定出对应于 nap、pol、ole 和 ca

结论：系统发育结果为芸苔属植物的细胞器基因组进化提供了新的见解。天然油菜籽至少含有 4 种细胞可塑性单

\*通信:huawei@oilcrops.cn; liujing@oilcrops.cn

刘洪芳和赵薇对这项工作做出了同样的贡献。

1 中国农业科学院油料作物研究所，农业和农村事务部油料作物生物和遗传改良重点实验室，武汉 430062。作者信息完整

列表见本文末尾



作者(s) 2022。开放访问本文是根据 Creative Commons Attribution 4.0 国际许可证获得许可的，该许可证允许以任何媒体或格式使用、共享、改编、分发和复制，只要您对原始作者和来源给予适当的信任，提供到 Creative Commons 许可证的链接，并指明是否进行了更改。本文中的图像或其他第三方材料包含在本文的 Creative Commons 许可证中，除非该材料的信用额度中另有说明。如果材料不包括在文章的创作共用许可中，并且您的预期用途不被法律法规允许或超过了允许的用途，您将需要直接获得版权所有者的许可。要查看此许可证的副本，请访问 <http://creativecommons.org/licenses/by/4.0/>。创作共用公共领域奉献豁免 (<http://creativecommons.org/publicdomain/zero/1.0/>) 适用于本文中提供的数据，除非在数据的信用额度中另有说明。

用于油菜细胞质遗传相关农艺重要性状的分析, 可为油菜品种的培育和改良提供实质性的便利。  
关键词: 芸苔属, 油菜, 细胞器全基因组, 有丝分裂型, 细胞质单倍型, 母系祖先

## 介绍

油菜籽是世界上最重要的油料作物之一, 也用作蛋白质饲料和工业原料来源[1]. 这是一个同种异体-大约 7, 500 年前起源于菜心(AA,  $2n = 20$ )和甘蓝(CC,  $2n = 18$ )杂交的倍性物种(AACC,  $2n = 38$ )[2-4]并最早在欧洲被种植[5]. 尽管它的关系-与其他作物如水稻相比, 驯化历史相对较短[6]和大豆(最大甘氨酸)[7], 油菜籽已经适应了不同的生态环境条件, 并根据生长和开花特性被驯化成三个不同的生态类型组, 即冬季、半冬季和春季类型[8-10]. 冬油菜和半冬油菜的春化时间分别大于 1 个月和 15 ~ 20 天, 而春油菜不需要春化。

为了更好地理解油菜表型多样性形成的进化和机制, 已经实施了大规模的基因组测序项目, 因为高质量的甘蓝型油菜基因组有助于在大量不同材料之间进行全基因组序列比较[3, 11-14]. 通过对来自全球 39 个国家的 991 份种质资源进行重新测序, 确定了油菜遗传多态性的全球格局, 进一步揭示了群体分裂和混合的途径, 揭示了生态型分化的遗传基础[15]. 另一个对 588 份甘蓝型油菜材料的重测序项目揭示了冬油菜种子可能是甘蓝型油菜的原始形态, 并通过整合全基因组关联研究、选择信号和转录组分析, 鉴定了与胁迫耐受性、含油量、种子质量和生态型改良相关的遗传位点[4]. 在收集 1688 份油菜重测序数据的基础上, 构建了由多组学数据和常用生物信息学工具组成的基因组平台 BnPIR 数据库, 该数据库包含基因信息、系统发育关系、表达数据和存在/缺失变异(pav)信息[16]. 群体重测序策略对于研究系统发育、系统地理学和群体遗传学信息尤其有效。

已对油菜籽基因组进行了广泛的重测序研究。然而, 鲜有研究集中在细胞器基因组上

遗传信息的重要组成部分。陆生植物的 chloroplast (cp)和线粒体(mt)基因组分别包含一个基因组大小相对较小的环状 DNA 分子: chloroplast 基因组的大小在 115 到 165 kb 之间[17], 线粒体基因组大小波动约 200 kb - 2mb[18]. 细胞质遗传对水稻的产量、耐低温性、粒重、充实粒率和碾磨品质等最重要的农艺性状的形成起着重要作用[19-21], 玉米株高[22], 大豆中的种子蛋白质含量[23, 24], 以及油菜籽中的含油量[25-27]. 到目前为止, 一些细胞器基因已经被证明在表型上是重要的。例如, 在我们之前的研究中, 线粒体编码的 orf188 被鉴定为一个潜在的菜籽油含量测定基因[28], 但细胞质活性决定农艺性状的遗传机制尚未得到充分探索。此外, 关键的结构和功能成分编码基因[29-31]对于更好地理解进化差异的机制至关重要[32-36]. 已经对 mt 和 cp 基因组进行了进化研究, 使用了同时、高效的提取和组装方法来获得质体和 mt 基因组[37], 为甘蓝型油菜的遗传研究提供了更深入的研究见解[38-42]. 然而, 大多数研究集中于在收集少量种质的基础上建立不同物种或品种之间的关系。近年来, 已在作物中进行了大量基于群体的 organelle 基因组研究。根据 412 个水稻 cp 和 mt 基因组分析, 发现籼稻和粳稻经历了不同的驯化过程[43, 44]. 甘蓝型油菜及其祖细胞的细胞器系统发育与细胞核相结合, 有助于揭示遗传和形成后渗入的不同模式[41]. 然而, 全球油菜细胞器基因组的遗传多样性、群体结构和农艺性状的遗传基础尚未得到深入分析。

在这项研究中, 我们开发了细胞器基因组数据-对大量不同的材料进行了全面的研究: 1, 579 份天然[4, 15]和 31 合成甘蓝型油菜[45]、199 B. rapa 和 119 B. 羽衣甘蓝- cea [46]. 细胞器基因组测序读数为

为每次加入提取和组装。接下来合并所有的环装配体，以构建细胞器泛基因组和等位基因变体数据集。我们的分析集中在系统发育关系、种群结构和遗传多样性上，尤其关注不同细胞质组的差异，这为油菜籽的起源和进化提供了新的细胞器基因组证据。此外，这些等位变异提供了全面的信息，可作为后续研究细胞质效应影响作物农艺性状的基础。此外，我们的发现有助于在不远的将来加速细胞器基因组辅助育种的进程。

## 结果

### 油菜籽细胞器基因组(cp和mt)的组装和注释

来自所有主要生产国的1,579份油菜籽样品的基因组测序数据来自SRP155312下的NCBI数据库[15]和PRJNA358784[4]。在对读取两端的低质量区域进行质量检查和修剪后，我们首先将干净数据分别映射到由35 mt基因组组成的线粒体基因组序列簇和由42 cp基因组组成的叶绿体基因组序列簇。这两个数据集都包含六种Brassica物种(表S1)。映射的成对端读取

对cp和mt基因组进行了提取和从头组装[47]和ARC软件(<http://ibest.github.io/ARC/>)，分别为。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( $> 100\times$ )。

细胞器基因组(cpDNA和mtDNA)是随后为每个入口单独组装。总计1,327个cpDNA和1,456个mtDNA被组装成单个圆形作图分子，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)。cp和mt基因组的主要G + C含量分别约为36.3%和45.2%

相应地(图1c, D)。在所有组装的序列中，47个cpDNAs和225个mtDNAs含有间隙填充碱基“N”，其长度在1至770 bp之间。

用GeSeq对组装的基因组进行了注释

[48]使用由从已知芸苔属物种获得的基因组组成的参考，所述已知芸苔属物种的基因组也用作上述mapping中的参考。去除重复项后，cp参考基因组包含79个蛋白编码基因、2个功能未知的开放阅读框(orf)、4个rRNAs和21个tRNAs。mt参考基因组包括35个蛋白编码基因、3个rRNAs、17个tRNAs和80个orf，其中33个

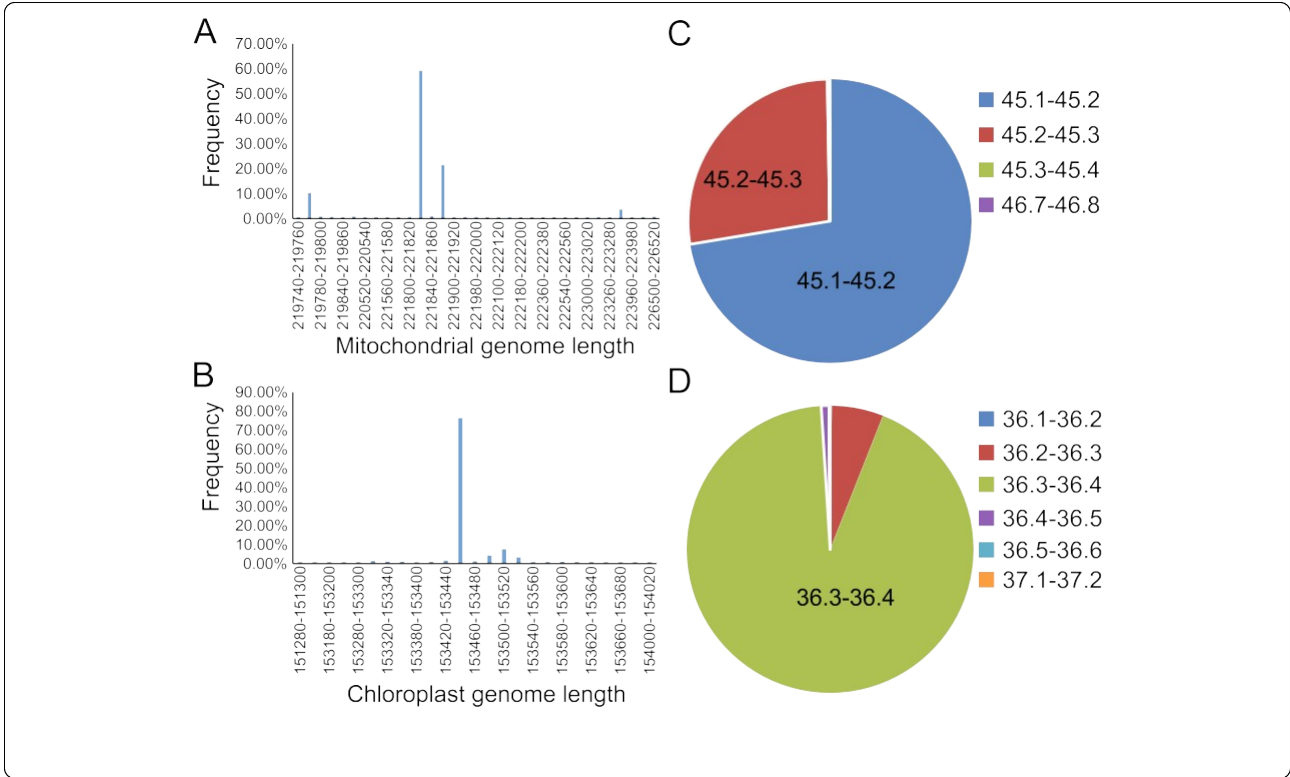


图1 组装叶绿体和线粒体基因组的序列特征。线粒体基因组a和B的长度分布  
叶绿体基因组。线粒体基因组和叶绿体基因组的C、G + C含量

其对应于不同有丝分裂类型不一致拥有的八个同源组[49]。

在 1, 327 个组装的油菜籽 cpDNAs 中的每一个中均检测到参考基因。由于多个基因存在多个拷贝, 99%的材料中鉴定出 87 个基因座对应的 79 个蛋白编码基因, 所有样本中鉴定出 8 个 rrna 基因座(所有 4 个 rrna 均为双拷贝), 99%的材料中鉴定出 37 个 tRNAs。在所有预测的叶绿体基因中, 95%的基因与参考基因具有超过 98%的相似性。

与 cp 相似, mt 基因也高度保守。除有丝分裂型特异性 orf(35 个蛋白编码基因、3 个 rRNAs、17 个 tRNAs 和 47 个 orf)外, 预测每个装配的基因组包含所有参考基因。基因序列与参考基因序列高度相似。在所有预测中, 99%的基因显示 98%的覆盖率, 93%的基因显示 98%的同一性。在后续分析中, 我们排除了 10%的基因与参考碱基对齐小于 80%的异常材料, 这远远低于其他样本。

如上所述, 所有的材料都被预测包含所有的参考基因, 这表明了基因含量方面的装配的完整性。为了进一步评估基因组组装的质量, 我们首先对公共芸苔属物种的细胞器基因组进行了比较分析。组装后的细胞器基因组通过 BLASTN 作图[50]过滤器 criteria 为 identity > 0.9, 且仅保留了每个样本具有最佳比对的受试者。集合与对应映射受试者之间的同源性序列超过 97.5%(所有查询和受试者的覆盖率> 97.5%), 表明当前集合已经完成。此外, 我们将用于组装的读取映射回每次加入时组装的单个循环序列。覆盖范围和深度是根据完全匹配的读数来测量的, 这意味着读数不匹配、缺失、insert-

以及软或硬剪辑都被过滤掉。97.7%的装配体被沿着整个基因组的至少 150 × 读数覆盖, 并且图谱的曲线图-

沿着基因组体的 ping 深度大致稳定(图 S1S, 2), 这保证了在单核苷酸水平的组装的高水平的准确性。此外, 为了评估我们的程序集中是否有结构错误, 我们检测了读取的连续性

绘制基因组上的起始位点。此处的间隙定义为 > = 150 bp, 无任何读取映射起点。如果检测到任何间隙, 可能会有断点, 因为大多数

读取长度为 150 bp, 序列深度大于 100 倍。我们发现, 在 80%的组件中未检测到间隙(无 “N” 序列)。这

结果表明, 80%以上的材料在结构水平上对独特的基因组区域进行了高质量的组装, 进一步证实了甘蓝型油菜细胞器基因组序列的高质量组装。

### 甘蓝型油菜细胞器全基因组的构建

我们使用基于参考的组装方法构建了甘蓝型油菜的细胞器全基因组(附加文件 1)。R4834 的细胞器基因组被视为 mt 和 cp 草案基因组, 因为它们与其他材料的比对数量最多。基于装配好的甘蓝型油菜 cp 和 mt 的全基因组比对并参考文献, 使用本地 Perl 脚本进行变体调用。通过插入插入片段(> 10 bp)编辑基因组草稿, 并通过添加至少有两个材料支持的未锚定片段(> 100 bp)进行补充。

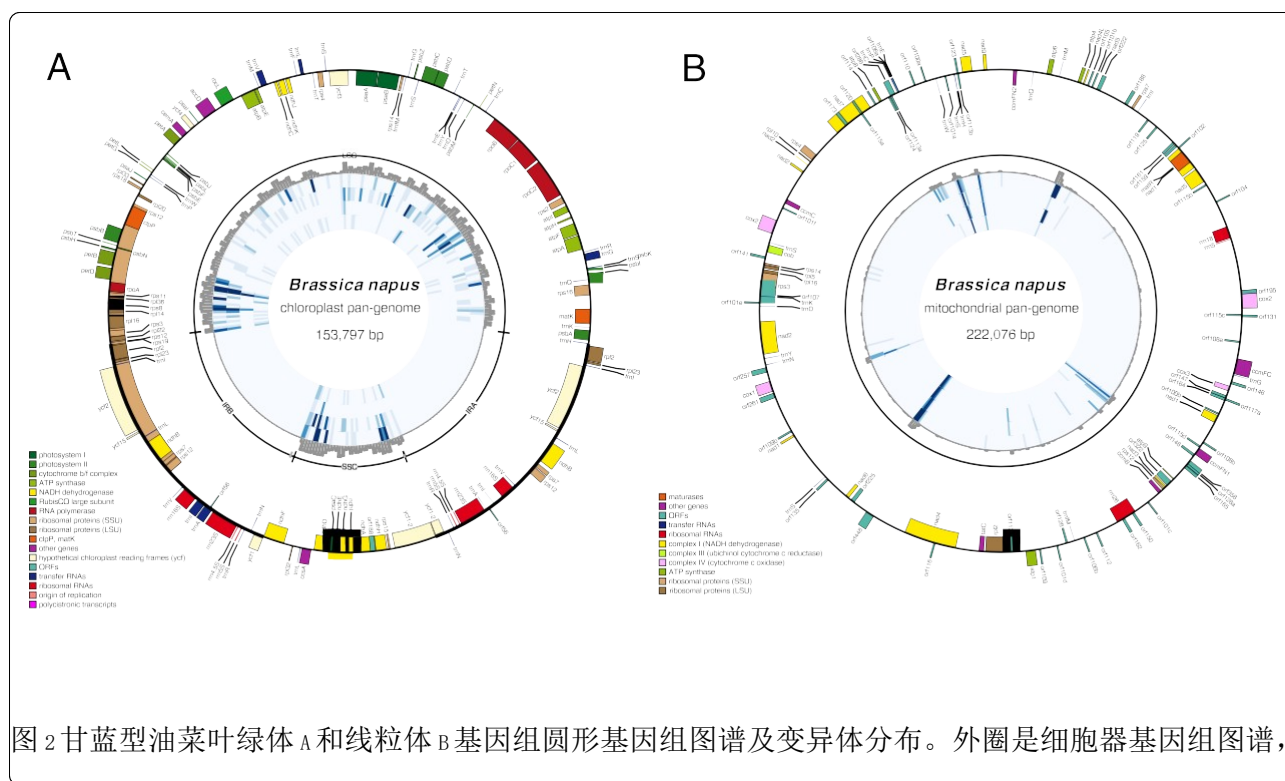
比对结果显示 cp 基因组结构稳定, 几乎没有重组序列。cp 泛基因组是一个 153, 797 bp 的环状分子, 由两个反向重复序列之间的一个大的和一个小的单拷贝(LSC 和 SSC)区域组成。我们共鉴定出 87 个蛋白, 3 个功能未知的 orf, 8 个核糖体 RNA 和 37 个 tRNA 序列(图 2a), 约占基因组的 49%。

mt pan 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 的可分配支架组成(图。2b)。共有 35 个蛋白编码基因,

注释了 3 个 rRNAs、18 个 tRNAs 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 [52]。

细胞器基因的有无变异(PAV)是影响细胞器遗传农艺性状的重要遗传因素。在此, 全基因组组装和基因注释为全球范围内甘蓝型油菜细胞器基因组的 PAV 分析提供了信息。我们将基因存在-缺失标准定义为覆盖率的 60%界限阈值和与参考值的 85%同一性界限阈值。所有编码蛋白的 cp 基因都是核心基因, 其中只有一个基因 ycf2 与仅一个片段中的一个额外拷贝不一致(表 S2)。所有四种核糖体基因在 1, 327 份材料中出现两次, 但 25 种 tRNAs 中有 3 种在群体中具有不同的拷贝数, 且在少数样品(< 1%)中均有所不同。总体而言, 在 1, 327 份材料中发现了所有 cp 基因, 但在少数材料中只有 4 份具有不同的拷贝数。





在不考虑 mt 基因组中有丝分裂型特异性 orf 的情况下, 在任何组装的附件中都没有缺失, 在不到 1% 的附件中, 仅发现 3 个 orf (orf 108 a、orf115c 和 orf131) 和 1 个 tRNA (trmH) 具有不同的拷贝数 (表 S3). 作为上述信息的总结, mt 基因组的所有蛋白编码基因都是核心基因, 其中 4 个的拷贝数不同。

已知 cp 基因组高度保守, 而在 mt 基因组中观察到由重排诱导的广泛重组和序列增益与损失[49, 53, 54]. 因此, mt 基因组中的序列变体占细胞质特异性序列的大多数。为了区分所有组合材料的细胞质类型, 亨等人开发了 12 个有丝分裂型特异性标记 (基因组特异性序列, > 100 BP)[55] 和 34 个同源 ORFs[49] 这在六种芸苔属有丝分裂类型中是不一致的, 用于搜索每个 mt 基因组。我们为样本定义了完全符合线粒体类型与 MSSs/orf 之间关系的 cello plas-mic 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“类似”类型。

[56]对齐。为了检查细胞器基因组的结构特征, 选择了来自不同细胞质的四个 cp 和四个 mt 基因组来构建多重基因组比对, 并通过 Mauve(<http://darlinglab.org/mauve/mauve.html>) (图 S3). 来自不同细胞质的 cp 基因组对整个基因组表现出共线性。然而, 这四种 mt 基因组有丝分裂却发生了重排。

表 1 1456 份甘蓝型油菜组合材料有丝分裂类型的分类和分布

<i>nap</i>	1.	215	83.45
<i>凸輪</i>	170	11.68	
<i>pol</i>	53	3.64	
<i>ole</i>	18	1.24	

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，我们分别建立了不同细胞质在三种生态型中的分布和不同生态型在四种细胞质中的分布(图 S4). 3 个生态型间细胞质分布无明显差异，与总体分布相似，但春季组 cam 型 28%例外，少于其他 3 组的 10%。然而，三种生态型在不同细胞质类型中的比例不一致。pol 型种质以半冬型为主(83.63%)，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细胞质变异

在组装的 cp 和 mt 基因组中分别检测到 4, 115 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 2, 092 个高质量变异体，包括 1, 660 个 SNPs (79.35%)和 433 个小 InDels (20.65%)。此外，在 mt 基因组中鉴定出 254 个 SNPs (77.91%)和 72 个小 indel (22.09%)，相应地在 cp 和 mt 基因组中达到每 kb 13 个和 1.5 个变异体的密度(表 2)。在整个群体中，cpD na 中检测到的 SNP 和小 indel 数量远远高于 mtDNAs 中检测到的数量，但每个亚组中 CPD na 中的数量较低，表明不同细胞质之间的 SNP/小 indel 在 CP 中比在 mt 中更丰富，而在 CPD na 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，其阅读被跳过(图 2)。在总共 1, 327 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(表 3 表, S5)，导致可能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-value < 0.0001, phyper 检验)。值得注意的是，45%的变体位于编码区，这比所有编码区 28%的基因组序列的频率高得多。在编码区的变体中，58%被预测为受功能影响(表 S5)。在四种细胞质类型中检测到的 cp 变异体的比例与 mt 不相似

表 3 变体(SNPs 和 InDels)的基因组分布

区域 Cp 变体 Mt 变体		
下游/上游 1, 279 147		
外显子		
同义/非框架移位 427		64
非同步/移码 322		76
stopgain 5		—
停用字词表 1		—
未知 0		10
基因间 0		26
内含子 369		8
拼接 0		—
ncRNA _外显子 4		—
总计 a	2, 407	335

a 这里的总数大于上述数字，因为几个基因座有多个变异体

表 2 总变体和亚组变体(SNPs 和 InDels)汇总

组	变体 Cp			Mt 变体		
	多态性	插入缺失	Ts/Tv a	多态性	插入缺失	Ts/Tv
分组						
疏忽	131	64	0.5	213	49	1.01
凸轮	104	75	0.62	90	24	0.92
政客	9	17	0.29	138	36	0.97
旧的	31	17	0.58	64	8	—
全部	1,660	433	0.78	254	72	0.95

a Ts/Tv 是转换/颠换的比例

基因组。此外，还观察到转换/转译(Ts/Tv)比率不一致(表2). 四种有丝分裂类型在 mt 基因组中的比例约为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) 中有四个常见 cp SNPs 和 57 个常见 mt SNPs。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)。

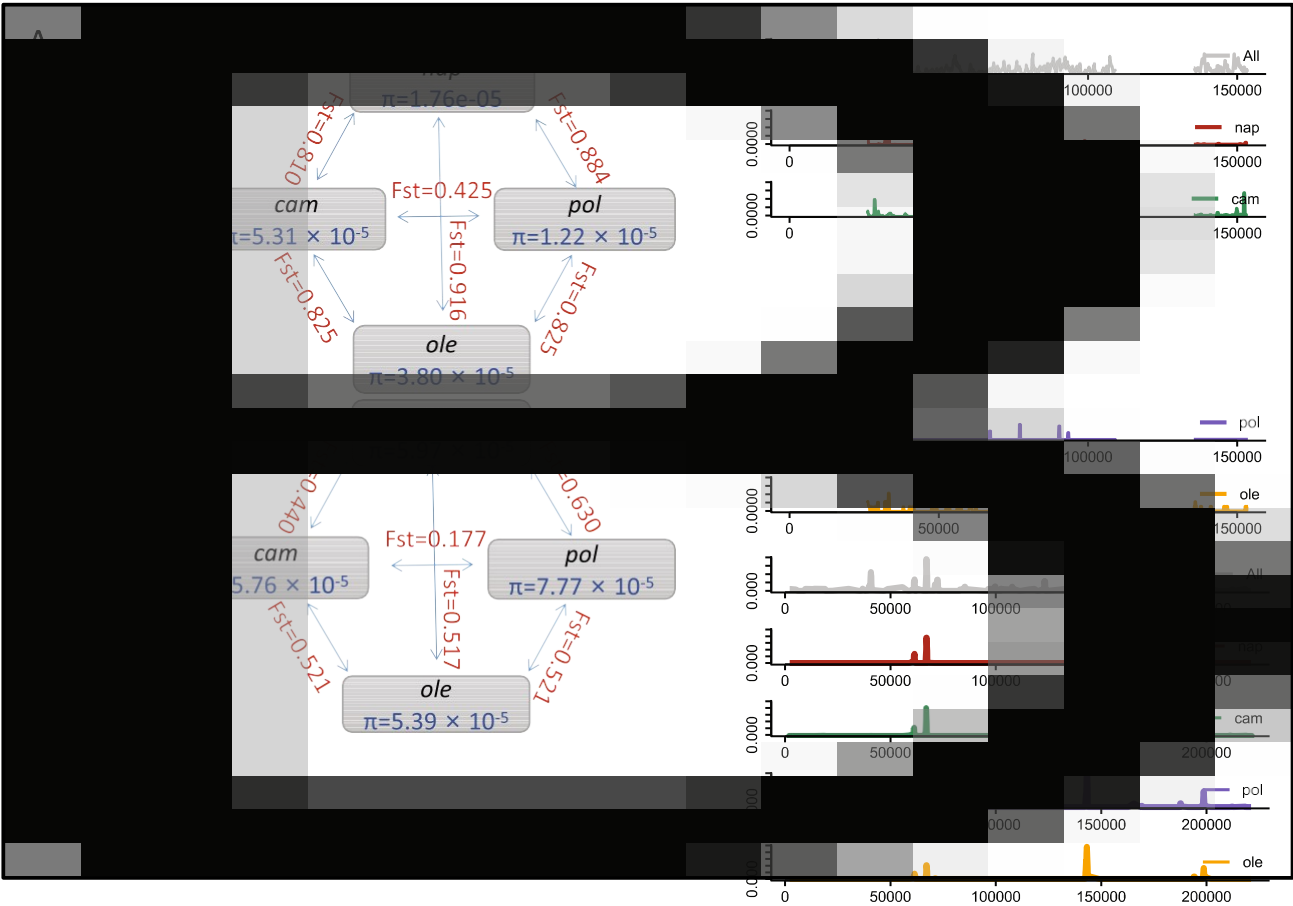
细胞质基因组的多样性分析

基于全基因组变体对 cp 和 mt 基因组进行了核苷酸多样性( $\pi$ )和固定指数(Fst)分析。我们发现 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。最高的潜水者—  
mt 基因组 pol 中的丰度为  $7.77 \times 10^{-5}$ ，cp 基因组 nap 中的低 est 为  $1.76 \times 10^{-5}$ (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  $6.68 \times 10^{-4}$  和  $1.63 \times 10^{-4}$  积极地。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 500 BP 滑动窗口和 100 BP 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  $P > 0.05$ )，而 cp 基因组间存在显著差异(t 检验  $P < 0.01$ )(图 1)。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，我们计算了多样性的减少



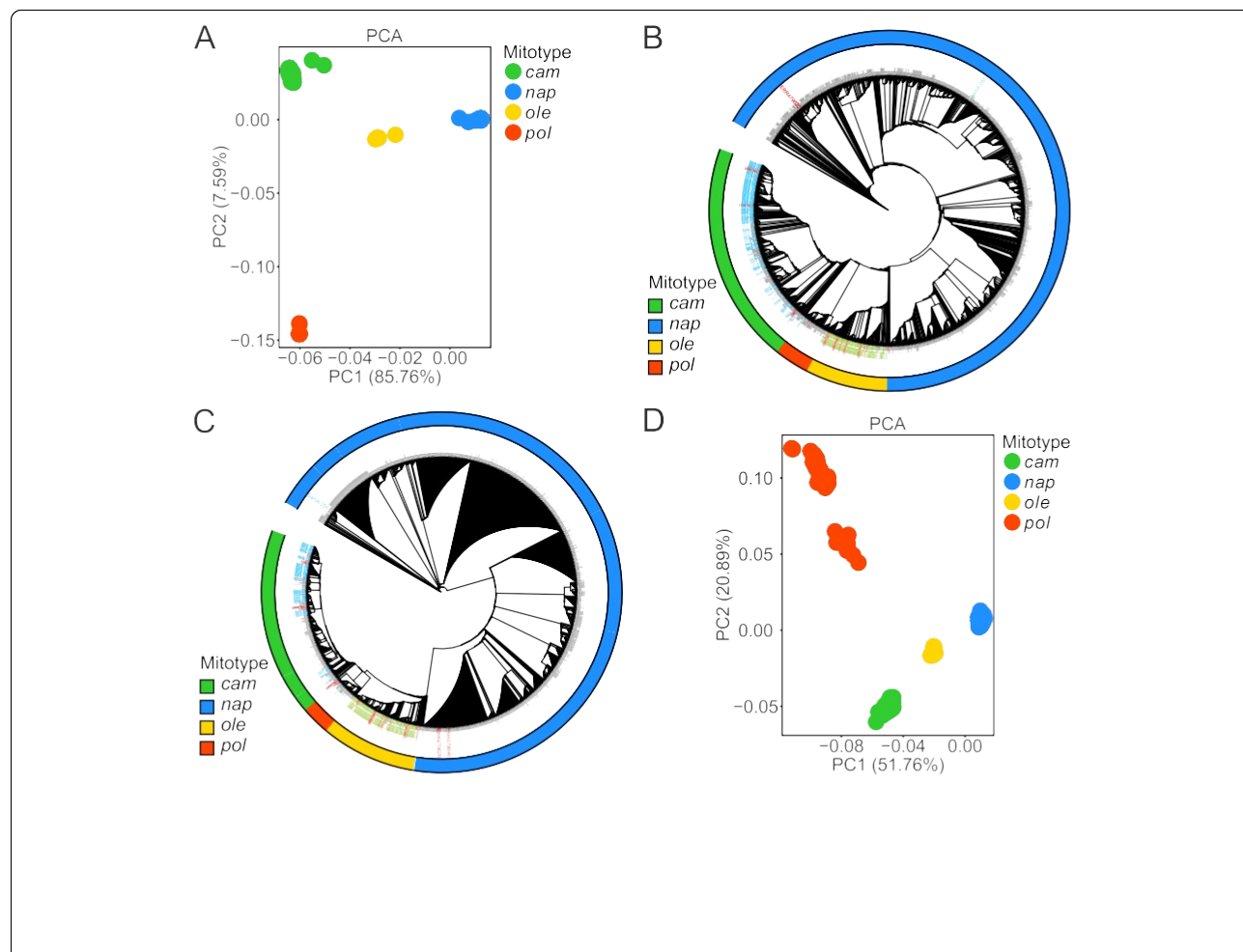


图 4 甘蓝型油菜材料及其两个祖种的群体结构。组装的  $cp$  基因组的 PCA 图。组合甘蓝型油菜与 199 份甘蓝型油菜、

表 4 不同细胞质间叶绿体基因组区域的显著差异

开始	目标	n_变体	$\pi$ (所有样本)	$\pi$ (nap)	$\pi$ (凸轮)	$\pi$ (pol)	$\pi$ (ole)	基因
26,001	26,800	16	3.27e-03	1.03e-03	1.95e-04	na a	3.44e-04	<i>trnH, psbA</i>
30,001	30,700	38	4.77e-03	1.36e-05	7.09e-05	7.04e-04	钠	<i>matK(up stran)</i>
33,301	34,200	29	4.10e-03	7.75e-04	1.46e-04	钠	2.82e-04	<i>psbI, trn</i>
55,401	56,100	32	3.34e-03	钠	钠	钠	钠	盗窃提货不着 险
81,501	82,100	19	3.79e-03	钠	1.32e-04	钠	钠	大亚基基因
88,301	88,800	38	3.54e-03	钠	2.64e-05	钠	钠	善待动物组织 (上游)
92,301	92,900	29	3.14e-03	9.50e-06	4.19e-04	钠	钠	<i>psaJ, rpl33</i>
107,301	107,800	15	3.94e-03	2.61e-05	7.76e-05	钠	钠	<i>rpl16</i>
138,901	139,600	34	3.25e-03	2.36e-05	钠	钠	钠	<i>rpl32</i>
148,901	149,600	18	3.26e-03	2.71e-06	钠	钠	钠	<i>rps15</i>
150,401	150,900	20	3.60e-03	钠	钠	钠	钠	<i>ycf1-2</i>

\*na\* 表示分组样本中没有 SNPs，但表示两个分组之间的几个变体



(ROD) [4] 基于使用 500-bp 窗口的整个组与 nap、cam、pol 和 ole 的核苷酸多样性比值的值(补充表 S6). 具有最高 1% ROD 值的区域被排除。连续窗口然后被合并成 11 个区域(表 4)。

### 遗传结构和系统发育关系

为了研究甘蓝型油菜及其二倍体亲本与人工合成的油菜细胞器基因组之间的遗传结构和系统发育关系, 对 199 份甘蓝型油菜和 119 份甘蓝型油菜材料进行了 SNPs 分析[46], 以及 31 份合成甘蓝型油菜材料[45] (11 份来自未发表的测序数据)通过对甘蓝型油菜细胞器泛基因组的定位读取进行检测。接下来, 在筛选出甘蓝型油菜中等位基因频率较小 ( $MAF < 0.02$ ), 但三个物种中  $MAF < 0.005$  的位点后, 我们进行了主成分分析 (PCA) 以及系统发育树和种群结构分析, 以保留物种特异性位点。基于 PCA 和最大似然 (ML) 树结果, 我们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的生态类型无关(补充图 S6)。来源于 cpDNA 和 mtDNA 的 PC1 的相关系数为 0.96(补充图 S7), 表明甘蓝型油菜自然群体中的 cp 和 mt 基因组同时进化, 因为 PC1 方差比例对于 cpDNA 为 85.76%, 对于 mtDNA 为 51.76%。

根据区分 nap 的群体结构结果, 确定甘蓝型油菜材料的最佳群体聚类数为  $K = 2$

和其他有丝分裂簇(补充图 S8)。然而, 在 PCA 图中, 通过 PC1 和 PC2, 甘蓝型油菜进一步聚类为 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 分组接近 PC1 水平。

为了研究三种 *Brassica* 种内和种间不同细胞质类型的驯化历史, 我们构建了一个包含甘蓝型油菜、甘蓝型油菜、合成和天然甘蓝型油菜群体的系谱的树。mt 和 cp 树显示了相似的种群结构。甘蓝型油菜材料分化为 3 个分支, 揭示了甘蓝型油菜的多倍体性(图 4b, C)。在先前的研究中, 少数甘蓝型油菜材料与大多数菜青虫归为一类[41]。在这里, 几乎所有的菜青虫都被归为最接近 cam 的一组, 其次是 pol of

*B. 甘蓝型油菜的 cam 和 pol 细胞质*

*B. 甘蓝型油菜可能是由菜青虫的祖先进化而来。此外, 所有的马齿苋种质都聚集在同一支中, 表明另一个独立的种质来源于马齿苋(图. 4b, C)。nap 有丝分裂型的配偶体, 在 naturalB 中占优势。甘蓝型油菜没有与甘蓝型油菜或甘蓝型油菜种质聚类在一起。但是, 加入了一个 *b. rapa* (SamA \_ 055, ssp. 和两个合成的甘蓝型油菜材料 (ERR479628 和 ERR479646), 它们的母本来自菜青虫。油茶*

*4 × 和 B. rapa ssp. 分别对菜青虫进行了聚类分析进入 nap clade, 而其他合成材料,*

*其母本为菜青虫的材料被聚类到 cam 分支中, 而其母本来自羽衣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, 支持聚类评估结果。上述信息表明, nap 细胞质可能是从雷帕氏杆菌的细胞质进化而来, 这在过去的研究中被认为是具有争议和不明确的。为了更清楚地研究不同类型细胞质之间的系统发育关系, 我们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)。*

### 讨论

MSS 标记和有丝分裂型特异性同源 ORFs 检测是鉴定细胞质类型的有效手段。在这项研究中。我们在 1456 份世界范围的甘蓝型油菜材料中鉴定了四种细胞质类型, 其中

18 份材料被归类为 ole, 其基因组包含相同的 MSSs 和 orf, 但缺少 ole 型中指定基因的双拷贝[49]。据报告, ole 型 (JF920286) mt 基因组是十字花科物种中最大的, 因为它有 141.8 kb 的重复片段[49], 这导致 18 个 orf 的基因数目不同。然而, 由于结构可变, 马齿苋 mt 基因组不同, 141.8 kb 片段不一定像马齿苋一样重复。葡萄孢[57]。由于重复序列在组装过程中更容易出错, 我们使用读取映射分析来检查 141.8 kb 重复序列的缺失是否是由 18 份 ole 型材料中的组装缺陷引起的。每次登录中用于程序集的读取被映射到相应的程序集, 并在删除重复项后检测到沿引用的深度。18 个等位基因的基因组深度曲线沿整个基因组分布均匀, 除了几个小区域 ( $< 3$  kb) 显示显著更高的深度(已证实为 cp 基因组的同源序列)和两个有丝分裂型特异性区域

发现了一些区域。与其他三组一致(补充数字 S1S, 2), 没有长片段(> 10 kb)具有双重覆盖深度, 揭示了甘蓝型油菜 ole 型 mt 基因组不包含重复的 141.8 kb 序列。

我们还通过将 MSSs 映射到重叠群来检测未组装成单个圆形分子的材料有丝分裂类型。少数材料被鉴定为具有不常见的有丝分裂类型。例如, R4699 [15] 具有 hau 细胞质, 其他两个基因组 (R4580, R5025 [4, 15]) 具有 ogu 细胞质。细胞质类型的鉴定对细胞质雄性不育系在杂种优势促进杂交育种中的广泛应用具有重要价值。

尽管已经进行了广泛的研究, 但油菜籽的起源和改良过程仍然不清楚和混乱, 因为还不知道真正的野生甘蓝型油菜群体。甘蓝型油菜是由甘蓝型油菜和甘蓝型油菜杂交而成, 其中 A 亚基因组被推测为来源于欧洲萝卜, C 亚基因组尚未确定 [4]. 近年来, 通过细胞器研究和分析, 对甘蓝型油菜的原始母系祖细胞进行了研究。在以往对甘蓝型油菜细胞器基因组的研究中, 一直认为 cam 型和 pol 型的起源是直接来自甘蓝型油菜遗传的, 但 nap 型存在争议且未确定。在系统发育树中, nap 型与葡萄型花椰菜紧密聚集在一起 [1], 因此认为它起源于花椰菜 [58]. 类似地, 推测某些稀疏 C-基因组野生种主要贡献了 nap 型细胞质, 并且是甘蓝型油菜的相应 C 亚基因组, 因为 nap 簇被插入 C-基因组分支的中间 [27]. 在我们的研究中还假设 nap 型在基于 mtDNA 的系统发育中最接近 ole 型 (图 S9). 此外, 由于其丰富的种质和 mt 基因组进化, nap 型甘蓝型油菜被进一步认为源自未鉴定或缺失的有丝分裂型油菜 [49]. 基于 chlo- roplast 和核遗传标记的组合, 在雷帕氏杆菌中以低频发现了最普遍的叶绿体单倍型, 但在 C 基因组材料 (野生和栽培的马齿苋及相关物种) 中不存在 [1], 这是我们细胞器重测序项目中的一致发现。在本研究中, 我们采集了大量甘蓝型油菜群体及其两个祖先物种 (菜豆和甘蓝) 的样本。利用 MSS 序列检测和系统发育分析, 首次在少数甘蓝型油菜和以甘蓝型油菜为母本的人工合成甘蓝型油菜中检测到优势 nap 型。因此, 我们推测午睡型 B. 甘蓝型油菜来源于一种罕见的有丝分裂型

有直接证据的雷帕菌。此外, 对 53 种 ole-型油菜的研究表明, 在天然油菜籽中也存在以 C 基因组物种为母体祖先的低频种质。

## 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甘蓝型油菜材料为基础, 构建了甘蓝型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。衍生自的系统发育树

B. 菜青虫, 羽衣甘蓝, 的天然和合成种群

B. 甘蓝型油菜揭示了不同甘蓝型油菜细胞质单倍型的起源。在甘蓝型油菜群体和母本为甘蓝型油菜的合成甘蓝型油菜中鉴定出 nap 型提供了新的论据, 表明甘蓝型油菜可能是 nap 型油菜的母体祖先。细胞质单倍型鉴定和系统发育为甘蓝型油菜细胞器基因组进化提供了新的视角。

## 材料和方法

### 样本和重新排序

本研究共利用两个已发表研究的 1579 份油菜种质构建了组织基因组和遗传多样性分析, 其中莱布尼茨植物遗传与作物研究所从 39 个国家收集了 991 份种质资源, 包括 3 个生态型 (658 个冬型、145 个半冬型和 188 个春型) (<https://gbis.ipk-gatersleben.de/gbis2i/faces/index.jsf>) 和浙江大学作物基因资源省级重点实验室 [15]. 另一个多样性小组由 588 份甘蓝型油菜材料组成 (74 份冬季, 428 份半冬季, 和

86 spring) 包括来自亚洲的 466 个、来自欧洲的 102 个、来自北美的 13 个和来自澳大利亚的 7 个 [4]. 此外, 还对 31 份人工合成的甘蓝型油菜、199 份甘蓝型油菜和 119 份甘蓝型油菜材料进行了系统发育关系的研究。199 份芜菁甘蓝和 119 份甘蓝材料来自植物育种种质、公司和基因库 [46]. 合成材料是通过高度多样性的亲本来源之间的种间杂交开发的, 其中 20 份材料是由 Schmutzer T, 等进行的 [45] 另有 11 个国家加入

由农业和农村事务部油料作物生物和遗传改良重点实验室测序。所有材料的全基因组 DNA 都是从叶子中提取的，并通过上述公共研究使用下一代测序 (NGS) 技术进行测序。补充表中列出了所有加入的详细信息 7-9。

### 泛基因组组装和注释

检查原始读取的质量，并通过修剪从读取的前沿和后沿修剪低质量区域 (版本 0.36，前沿:3 后沿:3 滑动窗口:4:15 MINLEN:120)

[59]。然后，使用 BWA 将干净的读数分别映射到线粒体基因组序列簇和叶绿体基因组序列簇 [60] 的 Sentieon DNaseq 软件 [61]。这两个数据集都包含构成 U [2] (表 S1)。对齐文件用于映射的成对端读取提取和从头组装。cp 基因组是使用 NOVOPlasty 版组装的 [47] CP 基因组和 mt 基因组的 kmer 值为 39，使用 ARC v 1.1.4-β (<http://ibest.github.io/ARC/>)。首先用 cover-age > 30 筛选由 ARC 组装的重叠群，并通过基于 BLAST 比对的本地 Perl 脚本连接到单个分子。

基于基因和基因组水平评估细胞器基因组装配的质量，基因注释揭示了装配在基因含量方面的完整性。然后，我们进行了基因组比较分析和对端阅读作图评估。使用 BLASTN 将组装后的基因组作图至上述参考 (公共芸苔属物种细胞器基因组序列)。筛选出同一性低于 90% 的比对，仅保留每个样本的最佳比对受试者。组装后的基因组与相应的参考基因组之间的高度相似性揭示了组装的完整性。使用 BWA 将用于组装的对端读取映射回组装的基因组 [60] 的 Sentieon DNaseq 软件 [61]。在过滤了不匹配、缺失、插入和软或硬剪切的读数后，我们通过测量覆盖范围和深度评估了组装的准确性，并检测了整个基因组中是否存在断点。

cp 和 mt pan 基因组是根据所有样本的全基因组比对构建的。首先，

通过 BLASTN (2.7.1+版) 将所有装配好的基因组相互对齐 [50] (-E 1e-30)，以及拥有数量最多的优质基因组

选择与其他样本的比对作为参考。所有其他程序集都映射到所选的

使用 Mugsy v1r2.3 参考 [56]。基于 Mugsy 比对，我们使用本地 Perl 脚本检测了插入片段 (> 10 bp) 和未锚定片段 (> 100 bp)。通过插入插入片段编辑基因组草稿，并以未锚定片段补充，这些片段被至少两个材料检测到。

为了探索甘蓝型油菜材料的基因和 PAV 信息 [48] (搜索标识 85) 被用于使用来自上述芸苔属物种的基因集作为参考注释细胞器泛基因组和所有组装样品的基因组。排除了覆盖率或同一性小于 60 的预测片段化基因。使用 OGDRAW 绘制了基因组图谱 [62] 和 Circos v 0.69-9 (<http://circos.ca/>)。

### 有丝分裂类型的鉴定

恒等开发的 12 个 MSS 标记 [55] 和 34 个同源 ORFs [49] 选择了 6 个不一致的芸苔属有丝分裂类型来区分所有组合材料的有丝分裂类型。我们通过 BLASTN 搜索了每个 MSS 标记相对于基因组装配的序列 [50]。同时，我们对照已知有丝分裂类型的已发表基因组搜索序列，并确定了每种 MSS 的过滤阈值，得出 MSS4 的相似性阈值为 90%，MSS9 为 85%，其他为 80%。利用 34 个同源 orf 的 PAV 信息，生成了材料与细胞质片段/orf 之间的关联数据，并确定了具有一致对应关系的材料的有丝分裂类型。

### 变体调用

用于装配的读数最初被映射到每个样本的装配基因组，选择没有任何错配的读数并映射到细胞器-lar 泛基因组。对于 mt 基因组未组装成单分子的材料，通过映射到与十字花科基因组相连的组装重叠群来填充读数。删除重复项后的 Bam 文件被带入变量调用 pro-

使用来自 Sentieon DNaseq 的单倍型进行访问 [61] 选项 -emit\_conf = 20，-call\_conf = 20。合成和天然的来自雷帕氏杆菌、羽衣甘蓝、甘蓝型油菜的变体

使用 BCFtools v1.3.1 合并各组 [63]。为了确定每个样本中所有未检测到的基因座的基因型，我们使用 samtools v1.3.1 [51] 和 bcftools，对于深度 (深度 < = 30) 和质量 (Q < 20) 较低的样本中的基因座，基因型设置为 N。

系统发育推断与种群结构

筛选出次要等位基因频率大于 5%且缺失数据小于 20%的 SNPs，用于群体结构和系统发育推断研究。使用 IQ-TREE v1.6.12 构建最大似然 (ML) 树 [64]。由 ModelFinder 确定的最佳模型 [65] 并指定 1000 个复制，这是为超快速引导选择的。这些树由在线工具交互式生命之树 (iTOL) v3 (<https://itol.embl.de>) [66]。GCTA v1.25 对甘蓝型油菜材料进行了主成分分析和 Fst 估计 [67]。人口结构由 STRU 结构 v2.3.4 推断 [68] 实现了基于模型的聚类方法。图是用 R 包龄 ggplot2 (<https://cran.r-project.org/web/packages/ggplot2/index.html>)。

核苷酸多样性

跨 3 个遗传聚类的核苷酸多样性 ( $\pi$ ) 和群体固定统计量 (Fst) 用 vcf tools v 0.1.13 计算 (<https://vcftools.github.io>) 使用 500-bp 滑动窗口和 100-bp 步长。多样性的减少 (ROD) [4] 计算基于 nap、cam 和 pol 之间的核苷酸多样性比率的值，以检测成对基因组的分化区域。

缩写

cp: 叶绿体; mt: 线粒体; 主成分分析法; MAF: 次要等位基因频率; MSS: 有丝分裂型特异性序列; PAV: 存在-不存在变化。

补充信息

在线版本包含补充材料，可从以下网址获取 <https://doi.org/10.1186/s12864-022-08573-x>。

承认

本文的数值计算是在中国农业科学院油料作物研究所超级计算中心的超级计算机系统上进行的。

作者贡献

所有作者都做出了重大贡献。本研究由 h-f. l.、J.L. 和 W.Z. 设计。h-f. l. 和 W.Z. 收集数据，进行大部分数据分析，并撰写本文。W.H. 和 J.L. 提供了研究设施。J.L. 参与了手稿的修订。W.H. 协调了这项研究。所有作者都已阅读并同意手稿的出版版本。

提供资金

本研究得到国家自然科学基金农业科技创新计划 (CAAS-ZDRW2019003) 的资助

国家基金 (31871664) 和红山实验室重大项目 (2021HSZD004)。作者声明，研究是在不存在任何可能被视为潜在利益冲突的商业或金融关系的情况下进行的。

数据和材料的可用性

1, 579 个天然甘蓝型油菜的原始基因组序列可在 GenBank 中的 SRP155312、PRJNA358784 和 PRJNA430009 下找到。生的 20 份合成甘蓝型油菜材料的测序数据可在欧洲核苷酸档案 (<https://www.ebi.ac.uk/ena/browser/home>) 在项目编号 PRJEB5974 和 PRJEB6069 下。雷帕氏杆菌的原始序列和马齿苋可在 NCBI 数据库的 bio project accentation prjna 312457 下找到。获得的用于细胞器基因组读取提取的六种芸苔属植物的已发表的 cp 和 mt 基因组用于补充数据表 S1。本研究期间生成的芸苔属植物叶绿体和线粒体基因组序列数据集可从门德尔利数据 (<https://doi.org/10.17632/9g7kxvgnyr.1>)。

声明

伦理批准和参与同意

不适用。

同意公布

不适用。

相互竞争的利益

作者声明没有利益冲突。

作者详细信息

<sup>1</sup> 中国农业科学院油料作物研究所。

农业和农村事务部油料作物生物与遗传改良重点实验室，武汉 430062。2 湖北红山实验室，武汉 430070。

收到: 2021 年 11 月 1 日接受: 2022 年 4 月 19 日  
Published online: 30 April 2022

参考

1. 艾伦德·CJ, GJ 国王。利用叶绿体和核分子标记研究甘蓝型油菜二倍体物种的起源。BMC Plant Biol. 2010;10:54.
2. 甘蓝型油菜基因组分析，特别是甘蓝型油菜的实验形成和独特的施肥方式。Jpn J Bot. 1935;7:389–452.
3. Chalhou B, Denoeud F, Liu S, Parkin IA, Tang H, Wang X, 等. 植物遗传学. 新石器时代后甘蓝型油菜基因组中的早期异源多倍体进化。科学。2014;345:950–3.
4. 陆 k, 魏 l, 李 x, 王 y, 吴杰, 刘 m, 等. 全基因组重测序揭示了甘蓝型油菜的起源及其改良所涉及的遗传位点。Nat Commun. 2019;10:1154.
5. 《十字花科的遗传学和基因组学》。纽约: 斯普林格; 2011. 第 585–96 页。
6. Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, 等. 对 50 份栽培稻和野生稻材料进行重测序以鉴定农艺学重要基因。Nat Biotechnol. 2011;30:105–11.
7. 对 302 份野生和栽培大豆材料重新测序鉴定了与大豆驯化和改良相关的基因。Nat Biotechnol. 2015;33:408–14.
8. Becker HC, Engqvist GM, Karlsson B. 基于等位酶和 RFLP 标记的油菜品种和再合成品系比较。定理应用遗传算法。1995;91:62–7.
9. Chen S, Nelson MN, Ghamkhar K, Fu T, Cowling WA. 来自相似起源的等位基因多样性的不同模式: 中国和澳大利亚的油菜案例。基因组。2008;51:1–10.
10. 钱 W, 萨斯 O, 孟 J, 李 m, 弗劳恩 M, 荣格 c. 甘蓝型油菜的杂种优势模式. I. 春与中国半冬杂交。定理应用遗传算法。2007;115:27–34.



11. 孙 f, 范 g, 胡 q, 周 y, 关 M, 童 c, 等. 甘蓝型油菜“ZS11”优质基因组揭示了半冬型的渐渗历史. 植物 J. 201792:452-68.
12. 宋, 关 Z, 胡杰, 郭 c, 杨 z, 王 s, 等. 8 个优质基因组揭示了甘蓝型油菜的全基因组结构和生态型分化. 纳特植物. 2020;6:34-45.
13. Chen X, Tong C, Zhang X, Song A, Hu M, Dong W, 等. 一个高质量的甘蓝型油菜基因组揭示了转座元件的扩增、亚基因组进化和抗病能力. 植物 Biotechnol J. 202119:615-30.
14. 邹杰, 毛磊, 邱杰, 王明, 贾立, 吴德, 等. 亚洲异源四倍体油菜幼苗的全基因组选择足迹和有害变异. 植物 Biotechnol J. 201917:1998-2010.
15. Wu D, Liang Z, Yan T, Xu Y, Xuan, Tang J, 等. 对世界油菜资源的全基因组测序揭示了生态型分化的遗传基础. 模型厂. 2019;12:30-43.
16. 宋, 刘 DX, 谢, 杨 Z, 郭 L, 刘 k, 等. BnPIR:甘蓝型油菜全基因组信息资源, 共 1689 份. 植物 Biotechnol J. 202119:412-4.
17. 陆地植物质体染色体的进化:基因含量, 基因顺序, 基因功能. Plant Mol Biol. 2011;76:273-97.
18. Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, 等. 水稻线粒体基因组的全序列:开花植物进化过程中频繁的 DNA 序列获得和丢失. Mol Genet Genomics. 2002;268:434-45.
19. 陶 d, 徐 p, 周杰, 邓 x, 李杰, 邓 w, 等. 细胞质对籼稻粒重和充实粒率的影响. BMC Genet. 2011;12:53.
20. 石 c, 朱杰. 籼稻碾磨品质性状的细胞质和母体效应的遗传分析. 种子科学技术. 1998;26:481-8.
21. 陶 d, 胡 f, 杨 j, 杨 g, 杨 y, 徐 p, 等. 细胞质和细胞质-细胞核的相互作用影响粳稻农艺性状. 幼发拉底卡. 2004;135:129-34.
22. Khehra AS, Bhalla SK. 细胞质对玉米数量性状的影响. 定理应用遗传算法. 1976;47:271-4.
23. 辛格·L·哈德利, HH. 大豆种子蛋白质含量的母体和细胞质效应. 作物科学. 1972;12:583-5.
24. 梁海, 王 s, 王 t, 张海, 赵 s, 张敏. 大豆异黄酮含量的胚、细胞质、母体效应及其环境互作的遗传分析 [l . Merr .]. 农业科学中国. 2007;6:1051-9.
25. 吴, 石 CH, 张 HZ. 甘蓝型油菜蛋白质含量的胚、细胞质和母体效应及其环境互作的遗传分析. 56:69-73.
26. 王 x, 刘 g, 杨 q, 华威, 刘杰, 王 h . 甘蓝型油菜含油量的遗传分析. 幼发拉底卡. 2009;173:17-24.
27. 华伟, 李, 詹 GM, 刘杰, 李杰, 王 XF, 等. 甘蓝型油菜种子含油量的母体控制:硅壁光合的作用. 植物 J. 201269:432-44.
28. 刘杰, 郝伟, 刘杰, 范 s, 赵伟, 邓露, 等. 一个新的嵌合线粒体基因赋予多倍体油菜种子含油量细胞质效应. 模型厂. 2019;12:582-96.
29. 艾伦·JF. 为什么叶绿体和线粒体含有基因组? Comp Funct Genomics. 2003;4:31-6.
30. 罗斯·RJ. 维持生命:维持植物的叶绿体、线粒体及其基因组. 耶鲁大学生物医学系. 2019;92:499-510.
31. Daniell H, Lin CS, Yu M, Chang. 叶绿体基因组:多样性、进化及其在基因工程中的应用. Genome Biol. 2016;17:134.
32. myszczynski K, Górski P, lipiko M, Sawicki J . 对裸腹蚤 (Jungermanniales) 细胞器基因组的测序揭示了进化稳定的苔类植物有丝分裂基因组的结构和基因顺序中的第一个例外. BMC Plant Biol. 2018;18:321.
33. 里巴罗拉 M, 福斯特 JT, 陈 AP, Williams AL, Rice DW, Liu X, 等. 蓖麻豆细胞器基因组测序及全球遗传多样性分析. PLoS One. 2011;6:e21743.
34. Hazkani-Covo E, Martin WF. 量化基因组进化和人类健康中独立细胞器 DNA 插入的数量. Genome Biol Evol. 2017;9:1190-203.
35. McManus HA, FuikovdK, Lewis PO, Lewis LA, KG. 细胞器系统发育学为绿藻科的系统学提供信息
- 水网藻科 (绿藻科) 并为绿藻生命树中质体基因组的复杂进化史提供线索. Am J Bot. 2018;105:315-29.
36. Wang X, Cheng F, Rohlsen D, Bi C, Wang C, Xu Y, 等. 园艺植物细胞器基因组组装方法及比较分析. 第 2018 号决议; 5:3.
37. 郝伟, 范 s, 华伟, 王海. 同时获得质体和线粒体基因组的有效提取和组装方法. PLoS ONE. 2014;9:e108291.
38. Seol, Kim K, Kang SH, Perumal S, Lee J, Kim CK. 两种芸苔属植物 (黑芸苔和甘蓝) 的完整叶绿素全基因组. 线粒体 DNA 部分, 2017; 28:167-8.
39. 乔杰, 张 x, 陈 B, 黄 F, 徐 k, 黄 Q, 等. 通过重测序比较细胞可塑性基因组:对农业上重要的芸苔属的遗传多样性和系统发育的见解. BMC 基因组公司. 2020;21:480.
40. 不同有丝分裂类型在甘蓝型油菜线粒体基因组中共存. 2011;6:e17662.
41. An H, Qi X, Gaynor ML, Hao Y, Gebken SC, Mabry ME, 等. 转录组和细胞器测序揭示了甘蓝型油菜异源四倍体的复杂起源和多样性. Nat Commun. 2019;10:2878.
42. 被子植物中线粒体和质体独立细胞质遗传的机制. j 植物资源 2010; 123:193-9.
43. 程 l, 南杰, 朱世华, 龙格那帕 P, 闵 MH, Cao Y, 等. 粳稻和籼稻叶绿体基因组差异选择的特征. 赖斯 (纽约). 2019;12:65.
44. 程 L, 金 KW, YJ 公园. 通过对栽培稻中粳稻和籼稻进行广泛的线粒体基因组分析, 获得了鉴定过程中选择事件的证据. Sci Rep. 20199:10846.
45. Schmutzer T, Samans B, Dyrzka E, Ulpinnis C, Weise S, Stengel D, 等. 来自模型异源多倍体植物甘蓝型油菜的物种全基因组序列和核苷酸多态性. Sci 数据. 2015;2:150072.
46. Cheng F, Sun R, Hou X, Zheng H, Zhang F, Zhang Y, 等. 亚基因组平行选择与芜菁甘蓝和甘蓝的形态类型多样性和趋同作物驯化有关. 纳特·杰特. 2016;48:1218-24.
47. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty:从全基因组数据从头组装细胞器基因组. 核酸第 2017 号决议; 45:18.
48. Tillich M, Lehwark P, 佩利泽 T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq -细胞器基因组的通用和精确注释. 核酸第 2017 号决议; 45:W6-11.
49. Chang S, Yang T, Du T, Huang Y, Chen J, Yan J, 等. 线粒体基因组测序有助于揭示芸苔属植物线粒体基因组形成的进化机制. BMC 基因组公司. 2011;12:497.
50. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST:架构和应用. BMC Bioinf. 2009;10:421.
51. Li H, Handsaker B, Wysoker A, Fennell T, 阮 J, Homer N, 等. 序列比对/图格式和 samtools . 生物信息学. 2009;25:2078-9.
52. L'Homme Y, Stahl, Li, Hameed A, Brown GG. 芸苔属 nap 细胞质雄性不育与含有与 pol CMS 相关 orf224 基因类似的嵌合基因的 mtDNA 区的表达有关. Curr Genet. 1997;31:325-35.
53. Li P, Zhang S, Li F, Zhang S, Zhang H, Wang X, 等. 叶绿体基因组的系统发育分析阐明了六个经济上重要的芸苔属物种之间的关系, 这些物种构成了 u. Front Plant Sci 的三角形. 2017;8:111.
54. 薛, 王勇, 陈明, 董 s, 邵, 刘勇. u 型三角形的母体遗传与芸苔属植物线粒体基因组的进化过程. 前线工厂科学. 2020;11:805.
55. Heng S, Chen F, Wei C, Hu K, Yang Z, Wen J, 等. 基于新开发的有丝分裂特异性标记的甘蓝型油菜分子标记辅助选择育种研究. 植物细胞研究 2017; 36:901-9.
56. Angiuoli SV, Salzberg SL. Mugsy:紧密相关的整个基因组的快速多重比对. 生物信息学. 2011;27:334-42.
57. Grewe F, Edger PP, Keren I, Sultan L, Pires JC, Ostersetzter-毕然 O, 等. 11 种芸苔属线粒体基因组的比较分析



- 甘蓝线粒体转录组。线粒体。2014;19(Pt B):135-43。
58. 杨杰, 刘 g, 赵 n, 陈 s, 刘 d, 马文, 等. 线粒体基因组比较分析揭示了芸苔属植物的进化重排机制. *plant Biol(stutg)*. 2016;18:527-36.
  59. Bolger AM, Lohse M, Usadel b. trim atorial:一种用于 Illu- mina 序列数据的灵活微调器。生物信息学。2014;30:2114-20.
  60. 使用 burrows- wheeler 变换实现快速、准确的短读数校准。生物信息学。2009;25:1754-60.
  61. Kendig KI, Baheti S, Bockol MA, Drucker TM, Hart SN, Heldenbrand JR, et al. sentian DNaseq variant calling workflow 展示了强大的计算性能和准确性。前线杰特。2019;10:736.
  - 62 Greiner S, Lehwark P, Bock r. organilargenomedraw(OGDRAW)ver-sion 1.3.1:用于细胞器基因组图形可视化的扩展工具包。核酸第 2019 号决议; 47:W59-64。
  63. Danecek P, McCarthy SA. BCF tols/csqa:单倍型感知变异序列。生物信息学。2017;33:2037-9.
  64. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE:一种快速有效的估计最大似然概率的随机算法。Mol Biol Evol. 2015;32:268-74.
  65. Kalyaanamoorthy S, Minh BQ, Wong T, von Haeseler A, Jermin LS. ModelFinder:用于精确系统发育估计的快速模型选择。Nat 方法。2017;14:587-9.
  66. Letunic I, Bork p. Interactive tree of life(iTOL)v3:用于显示和注释系统发育树和其他树的在线工具。核酸第 2016 号决议; 44:W242-5。
  67. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA:全基因组复杂性状分析的工具。Am J Hum Genet. 2011;88:76-82.
  68. Evanno G, Regnaut S, Goudet J. 使用软件 STRUCTURE 检测个体集群数量:一项模拟研究。Mol Ecol. 2005;14:2611-20.

## 出版商须知

在出版的地图和机构附属关系中, 斯普林格自然基金会会对管辖权主张保持中立。

