[](http://crossmark.crossref.org/dialog/?doi=10.1186/s12864-022-08573-x%26domain=pdf)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

以及油菜叶绿体和线粒体的进化起源

Honfang Liu1，Wei Zhao1，Wei Hua1，2\*和Jing Liu1，2\*

**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是未知的，这为油菜系统发育研究提供了独特的材料。在此，基于全球收集的1579份油菜材料的从头组装，我们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,白花甘蓝的遗传多样性、系统发育关系.

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cam有丝分裂类型的四种主要细胞质单倍型，其中叶绿体基因组的结构比线粒体基因组更保守，没有任何重排。在叶绿体基因组中共检测到2，092个变异，而在线粒体基因组中仅检测到326个变异体，这表明叶绿体基因组的单碱基多态性水平高于线粒体基因组。基于全基因组变异多样性分析，在叶绿体基因组上鉴定出不同细胞质单倍型间的11个遗传差异区域。包含青菜、白花甘蓝、自然和人为油菜群体的系统发育树揭示了油菜细胞质的多种起源。cam型和pol型均来源于青菜，而ole型来源于白花甘蓝。值得注意的是，nap型细胞质在青菜群体和人为油菜中均有发现，这表明青菜可能是nap型油菜的原始祖先。

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种细胞质单倍型，其中主要的nap型可能来源于青菜。此外，细胞器泛基因组和总体变异数据提供了有用的资源

\*通信:huawei@oilcrops.cn；[liujing@oilcrops.cn](mailto: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](#_bookmark10)].这是一个大约7，500年前起源于青菜(AA，2n = 20)和白花甘蓝(CC，2n = 18)杂交的异源多倍体物种(AACC，2n = 38 )[[2](#_bookmark11)–[4](#_bookmark13)]并最早在欧洲被种植[[5](#_bookmark14)].尽管与其他作物如水稻和大豆[[7](#_bookmark17)]相比，它的驯化历史相对较短[[6](#_bookmark16)]，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](#_bookmark18)–[10](#_bookmark19)].冬油菜和半冬油菜的春化时间分别是大于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](#_bookmark12),[11](#_bookmark20)–[14](#_bookmark21)].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](#_bookmark22)].另一个对588份油菜材料的重测序项目揭示了冬油菜种子可能是油菜的原始形态，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](#_bookmark13)].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](#_bookmark23)].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chlo- roplast (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:chlo- roplast基因组的大小在115到165 kb之间[[17](#_bookmark24)]，线粒体基因组大小波动约200 kb–2mb[[18](#_bookmark25)].细胞质遗传对籼稻的产量、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](#_bookmark26)–[21](#_bookmark27)]，玉米株高[[22](#_bookmark28)]，大豆中的种子蛋白质含量[[23](#_bookmark29),[24](#_bookmark30)]，以及油菜籽中的含油量[[25](#_bookmark31)–[27](#_bookmark32)].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](#_bookmark33)]，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](#_bookmark34)–[31](#_bookmark35)]对于更好地理解进化差异的机制至关重要[[32](#_bookmark36)–[36](#_bookmark37)].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](#_bookmark38)]，为油菜的遗传研究提供了更深入的研究见解[[38](#_bookmark39)–[42](#_bookmark42)].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细胞器基因组研究。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](#_bookmark43),[44](#_bookmark44)].油菜及其祖细胞的细胞器系统发育与细胞核相结合，有助于揭示遗传和基因渗入形成的不同模式[[41](#_bookmark41)].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](#_bookmark13),[15](#_bookmark22)]和31份合成油菜[[45](#_bookmark45)]、199 青菜和119 白花甘蓝 .细胞器基因组测序reads被

每次提取和组装。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的NCBI数据库[[15](#_bookmark22)]SRP155312下和PRJNA358784 [[4](#_bookmark13)].在对reads两端的低质量区域进行质量检查和修剪后，我们首先将clean data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芸薹属物种(表S[1](#_bookmark9)).映射的双端reads分别

用NOVOPlasty ARC软件进行了提取和从头组装成cp和mt基因组[[47](#_bookmark47)]。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)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[1](#_bookmark0)a，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[1](#_bookmark0)c，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](#_bookmark48)]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参考基因集包含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参考基因集包括35个蛋白编码基因、3个rRNAs、17个tRNAs和80个orf，其中33个

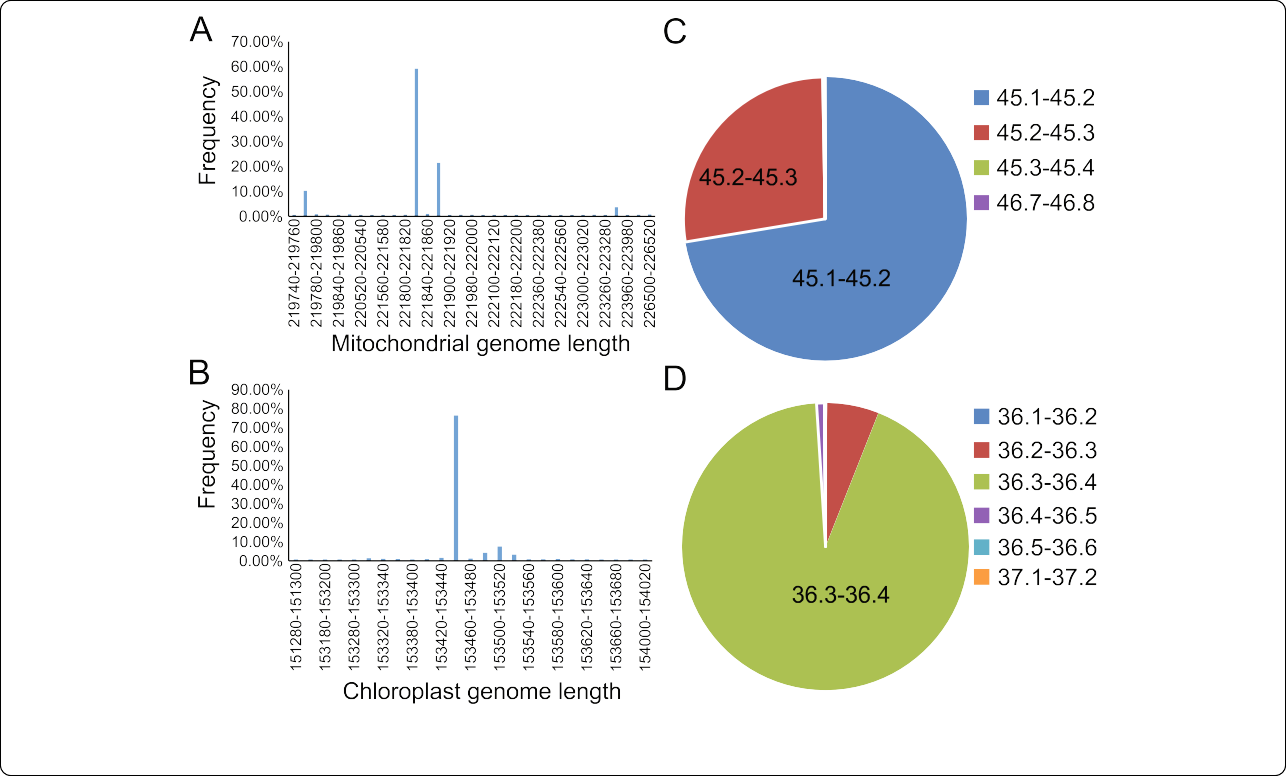


图1组装叶绿体和线粒体基因组的序列特征。线粒体基因组a和B的长度分布

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拥有不一致的八个同源组[[49](#_bookmark49)].

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%的一致性。在后续分析中，我们排除了与参考碱基对齐小于80%碱基的10%基因，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](#_bookmark50)](identity > 0.9)，且仅保留了每个样本具有最佳比对的结果。集合与对应映射对象之间的同源性序列超过97.5%(所有查询和subject的覆盖率> 97.5%)，表明当前集合已经完成。此外，我们每次将用于组装的reads映射回组装的单个循环序列。覆盖范围和深度是根据完全匹配的reads来测量的，这意味着reads不匹配、缺失、插入

、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 ×reads覆盖，并且图谱的曲线图-

沿着基因组体的映射深度大致稳定(图S[1](#_bookmark15)S，[2](#_bookmark15))，这保证了在单核苷酸水平的组装准确的高水平。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

定位基因组起始位点reads的连续性。此处的间隙定义为> = 150 bp，就是无任何reads映射起点。如果检测到间隙，这可能会有断点，因为大多数

reads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，进一步证实了油菜细胞器基因组序列的高质量组装。

**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油菜的细胞器全基因组(附加文件[1](#_bookmark8)).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。基于装配好的油菜cp和mt的与参考全基因组比对，使用本地Perl脚本进行变体识别。通过插入插入片段(> 10 bp)来编辑基因组草稿，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[2](#_bookmark1)a)，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。[2](#_bookmark1)b)。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](#_bookmark52)].

细胞器基因存在-缺失变异(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60%覆盖率的界限阈值和85%与参考一致性的界限阈值。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在一个材料中与1个额外拷贝不一致(表S[2](#_bookmark9)).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%)中都有所不同。总体而言，在1，327份材料中发现了所有cp基因，只有4个基因在少数材料中具有不同的拷贝数。

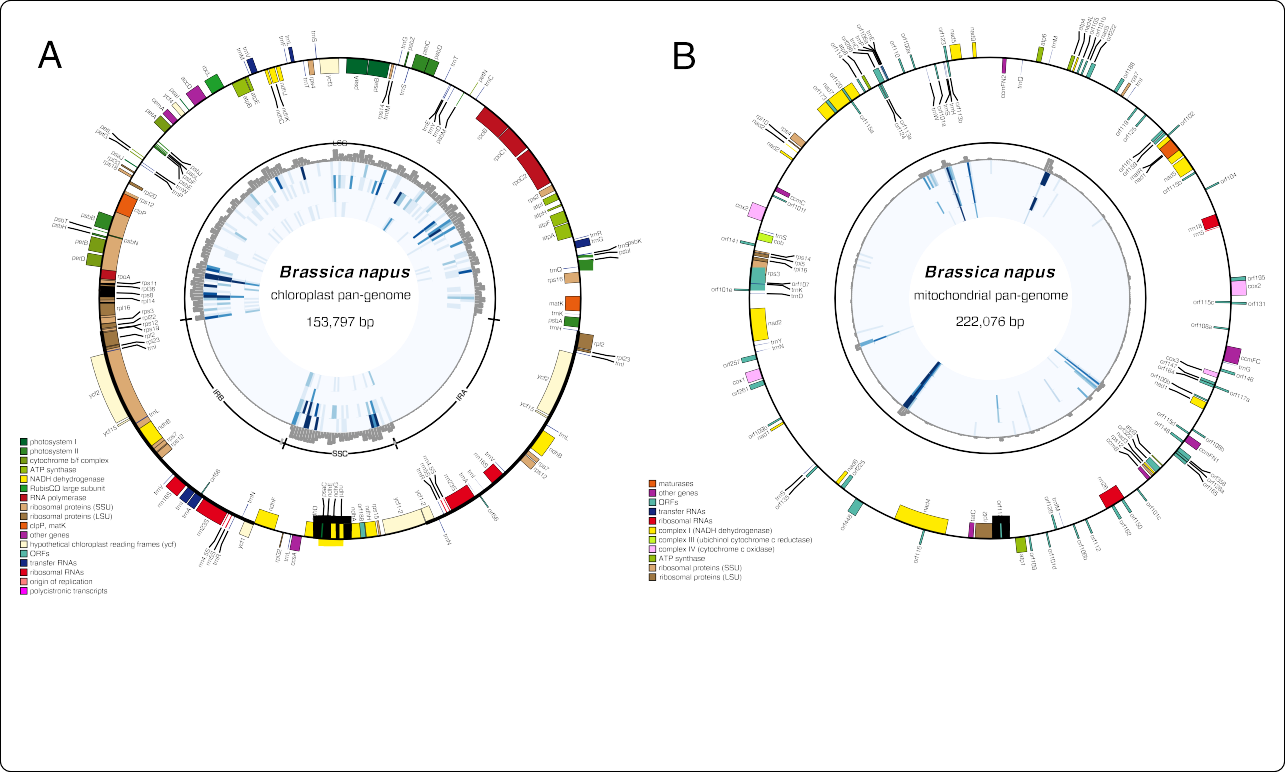


图2油菜叶绿体A和线粒体B基因组圆形基因组图谱及变异体分布。外圈是细胞器基因组图谱，显示了以颜色区分的不同功能类别(列于图例中)的基因。标记在圆圈内的基因是顺时针转录的，而圆圈外的基因是逆时针转录的。内圈显示了不同组中变体的分布，深灰色条代表所有组合的材料，蓝色高亮显示对应于四种不同的细胞质(nap、cam、pol和ole)，条越高或颜色越深，密度越大。每个条形图的长度表示500 BP窗口中的变体总数

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不存在缺失，仅发现3个orf(orf 108 a、orf115c和orf131)和1个tRNA (trmH)在不到1%的材料中具有不同的拷贝数(表S[3](#_bookmark9)).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。

**细胞质单倍型的鉴定**

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获得与缺失[[49](#_bookmark49),[53](#_bookmark53),[54](#_bookmark54)].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](#_bookmark55)]和34个同源ORFs[49](#_bookmark49)]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。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1，456份组合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](#_bookmark2)补充，表S[4](#_bookmark9)).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[[39](#_bookmark40)].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

[[56](#_bookmark56)]。为了检查细胞器基因组的结构特征，使用Mauve(<http://darlinglab.org/mauve/mauve.html>)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鉴定大规模的重排(图S[3](#_bookmark15)).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。

表1 1456份油菜组合材料有丝分裂类型的分类和分布

**细胞质类型数百分比(%)**

*nap 1，215 83.45*

*凸轮170 11.68*

*pol 53 3.64*

*ole 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S[4](#_bookmark15)).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](#_bookmark4)).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，但在CPD na中其每个亚组中的变异体数量较低，表明不同有丝分裂型之间的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reads被跳过(图[2](#_bookmark1)).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](#_bookmark3),表S[5](#_bookmark9))，会导致一个潜在的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，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(表S[5](#_bookmark9)).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 | 一 |
| 停用字词表1 | 一 |
| 未知0 | 10 |
| 基因间0 | 26 |
| 内含子369 | 8 |
| 拼接0 | 一 |
| ncRNA \_外显子4 | 一 |
| 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 | 一 |
| 全部 | 1,660 | 433 | 0.78 |  | 254 | 72 | 0.95 |

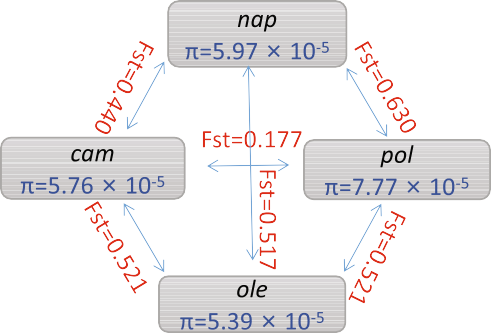
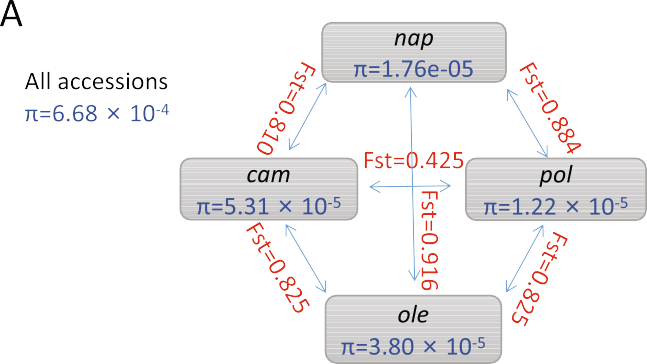
a Ts/Tv是转换/颠换的比例

基因组。此外，还观察到转换/颠换(Ts/Tv)比率不一致(表[2](#_bookmark4)).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S[5](#_bookmark15)).

**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。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[3](#_bookmark5)a，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[3](#_bookmark5)a、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[4](#_bookmark6)a，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[3](#_bookmark5)c，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3油菜四种主要细胞质类型期间细胞器基因组和亚基因组范围的核苷酸多样性和FST比较。

a基于cpDNA的四组之间的核苷酸多样性和Fst。b基于mtDNA的四组之间的核苷酸多样性和Fst。c除两个反向重复区外，cp基因组核苷酸多样性。d沿着mt基因组的核苷酸多样性。沿着不同细胞质组的细胞器基因组，对于具有100-bp步长的500-bp滑动窗口估计π值

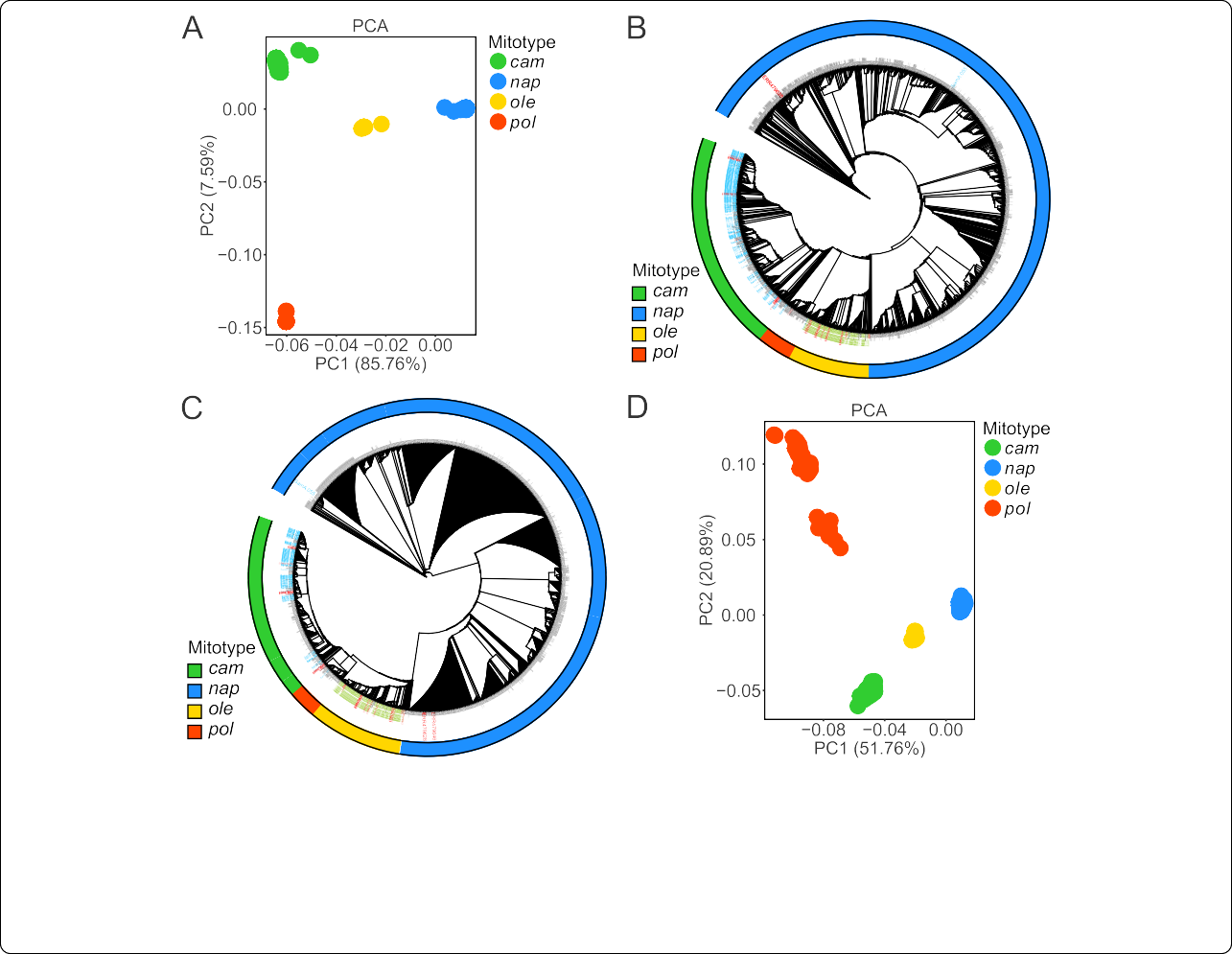


图4油菜材料及其两个祖种的群体结构。组装的cp基因组的PCA图。组合油菜与199份油菜、119份油葵和31份人工合成油菜材料的叶绿体系统发育。c .油菜及其两个祖种的线粒体系统发育。d组装mt基因组的PCA图。在筛选次要等位基因频率(MAF)< 0.005和缺失呼叫> 20%的位点后，使用SNPs构建ML系统发育树。在PCA图中，不同细胞质类型的样本以不同颜色的点突出显示。在系统发育树中，每个样本的细胞质类型在外圈标注为以不同颜色突出显示的vetical bar。系统发育树中每个样本的标签也用对应于以下组的其他四种颜色突出显示:黑色，天然油菜；红色合成油菜；黄绿色，B. olracea青色，B. rapa

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** | **π(所有样本)** | **π (nap)** | **π(凸轮)** | **π (pol)** | **π (ole)** | **基因** |
| 26,001 | 26,800 | 16 | 3.27e-03 | 1.03e-03 | 1.95e-04 | na a | 3.44e-04 | *trnH，psbA* |
| 30,001 | 30,700 | 38 | 4.77e-03 | 1.36e-05 | 7.09e-05 | 7.04e-04 | 钠 | *matK(up stran)* |
| 33,301 | 34,200 | 29 | 4.10e-03 | 7.75e-04 | 1.46e-04 | 钠 | 2.82e-04 | *psbI，trn* |
| 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 | *psaJ，rpl33* |
| 107,301 | 107,800 | 15 | 3.94e-03 | 2.61e-05 | 7.76e-05 | 钠 | 钠 | *rpl16* |
| 138,901 | 139,600 | 34 | 3.25e-03 | 2.36e-05 | 钠 | 钠 | 钠 | *rpl32* |
| 148,901 | 149,600 | 18 | 3.26e-03 | 2.71e-06 | 钠 | 钠 | 钠 | *rps15* |
| 150,401 | 150,900 | 20 | 3.60e-03 | 钠 | 钠 | 钠 | 钠 | *ycf1-2* |

“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](#_bookmark13)]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500-bp窗口(补充表S[6](#_bookmark9)).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](#_bookmark7)).

**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合成的油菜籽细胞器基因组之间的遗传结构和系统发育关系，通过对油菜细胞器泛基因组的定位reads进行检测199份青菜和119份白花甘蓝材料[[46](#_bookmark46)]，以及31份合成油菜材料[[45](#_bookmark45)] (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，但在保留物种特异性位点的三个物种中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。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导致的生态分型无关(补充图S[6](#_bookmark15)).来源于cpDNA和mtDNA的相关系数PC1为0.96(补充图S[7](#_bookmark15))，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

(补充图S[8](#_bookmark15)).然而，在PCA图中，通过PC1和PC2，油菜进一步聚类为4组(图[4](#_bookmark6)a，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，我们构建了一个包含青菜、白花甘蓝、合成和天然油菜群体的系谱的树。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[4](#_bookmark6)b，C)。在先前的研究中，少数油菜材料与大多数青菜归为一类[[41](#_bookmark41)].在这里，几乎所有的青菜都被归为最接近油菜cam的一组，其次是油菜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*[4](#_bookmark6)b，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)和两个合成的油菜材料(ERR479628和ERR479646)，它们的母本来自青菜。油茶4 ×和B. rapa ssp青菜分别被归到nap，而其他母本为青菜的合成材料被聚类到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，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S[9](#_bookmark15)).

# 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型特有基因的双拷贝[[49](#_bookmark49)].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](#_bookmark49)]，这导致18个orf的基因数目不同。然而，由于结构可变，白花甘蓝mt基因组是不同的情况，141.8 kb片段不一定是必要的重复,比如像B. oleracea var. Botrytis [[57](#_bookmark57)].由于重复序列在组装过程中更容易出错，我们使用reads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。每次用于组装的reads被映射到相应的数据集，并在删除重复项后检测到沿参考序列的深度。18个ole材料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)以及发现两个有丝分裂型特异性区域

与其他三组一致的是(补充图S[1](#_bookmark15)S[2](#_bookmark15))，没有长片段(> 10 kb)具有双重覆盖深度，揭示了油菜ole型mt基因组不包含141.8 kb重复序列。

我们还通过将MSSs映射到contig来检测未组装成环状分子的材料的有丝分裂类型。少数材料被鉴定为具有不常见的有丝分裂类型。例如，R4699

[[15](#_bookmark22)]具有hau细胞质，其他两个基因组(R4580，R5025 [[4](#_bookmark13),[15](#_bookmark22)])具有ogu细胞质。细胞质类型的鉴定对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油菜群体。油菜是由青菜和白花甘蓝杂交而成，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](#_bookmark13)].近年来，通过细胞器研究和分析，对油菜的原始母系祖细胞进行了研究。在以往对油菜细胞器基因组的研究中，一直认为cam型和pol型的起源是直接从青菜遗传的，但nap型存在争议且未确定。在系统发育树中，nap型与葡萄孢型白花甘蓝紧密聚集在一起,因此认为它起源于白花甘蓝[[58](#_bookmark58)].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，并且对应于油菜的C亚基因组，因为nap簇被插入C-基因组分支的中间[[27](#_bookmark32)].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S[9](#_bookmark15)).此外，由于其丰富的种质和mt基因组进化，nap型油菜被进一步认为源自未鉴定或缺失的青菜有丝分裂型[[49](#_bookmark49)].基于叶绿体和核遗传标记的组合，在青菜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(野生和栽培的白花甘蓝及相关物种)中不存在[[1](#_bookmark10)]，这是我们细胞器重测序项目中的一致发现。在本研究中，我们采集了大量油菜群体及其两个祖先物种(青菜和白花甘蓝)的样本。利用MSS序列检测和系统发育分析，首次在少数青菜和以青菜为母本的人工合成油菜中检测到优势nap型。因此，我们推测nap型油菜来源于一种青菜中罕见的有丝分裂型

。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。

# 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，其中德国莱布尼茨植物遗传与作物研究所和浙江大学作物基因资源省级重点实验室从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) [[15](#_bookmark22)].另一个多样性小组由588份油菜材料组成(74份冬季，428份半冬季,86 春季)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](#_bookmark13)].此外，还对31份人工合成的油菜、199份青菜和119份白花甘蓝材料进行了系统发育关系的研究。199份青菜和119份白花甘蓝材料来自植物育种种质资源、公司和genebak数据库[[46](#_bookmark46)].人工材料是通过高度多样性的亲本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](#_bookmark45)],另有11个材料

由农业农村部油料作物生物学与遗传育种重点实验室测序。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7–9中列出了所有材料的详细信息。

**泛基因组组装和注释**

检查原始reads的质量，并用Trimmomatic (版本0.36，前沿:3后沿:3滑动窗口:4:15 MINLEN:120)从reads的前沿和后沿修剪低质量区域[[59](#_bookmark59)].然后，使用DNASeq软件[61]中的BWA将clean reads分别映射到线粒体基因组序列簇和叶绿体基因组序列簇.这两个数据集都包含构成Triangle of U [[2](#_bookmark11)](表S[1](#_bookmark9))的6种芸薹属物种.对齐文件用于映射的双端reads提取和从头组装。cp基因组是使用NOVOPlasty kmer值为39组装的[[47](#_bookmark47)], mt基因组使用ARC组装.首先筛选覆盖度> 30由ARC组装的contig，并通过基于BLAST比对的本地Perl脚本连接成单独分子。

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双端reads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)。去掉一致性低于90%的比对，仅保留每个样本的最佳比对subject。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Sentieon DNASeq软件[61]中的BWA将用于组装的双端reads映射回组装基因组[[60](#_bookmark60)].在过滤了不匹配、缺失、插入和软或硬剪切的reads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泛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](#_bookmark50)] (-E 1e-30)，然后选择与其他样本相比拥有数量最多的优质基因组作为参考。使用Mugsy v1r2.3将其他组装结果都映射到所选的

参考[[56](#_bookmark56)].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)。通过插入这些插入片段编辑基因组草稿，并以被至少两个材料检测到的未锚定片段作为补充草稿。

为了探索油菜材料的基因和PAV信息,使用GeSeq [48] (search identity 85)来注释细胞器泛基因组和所有组装样品的基因组,使用来自上述芸苔属物种的基因集作为参考。排除了覆盖率或一致性小于60的预测的片段化基因。使用OGDRAW[62]和Circos v0.69–9来绘制基因组图谱.

**有丝分裂类型的鉴定**

heng等开发的12个MSS标记[[55](#_bookmark55)]和在6个芸苔属有丝分裂类中不一致的34个同源ORFs[49](#_bookmark49)]被选择用来区分所有组装材料的有丝分裂类型。我们通过BLASTN搜索了每个MSS标记相对于装配基因组的序列[[50](#_bookmark50)].同时，我们搜索了已知有丝分裂类型相对于已发表基因组的序列，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。

**变体识别**

对于每个样品,用于装配的reads先被映射到装配基因组，选择没有任何错配的reads并映射到细胞器泛基因组。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contig来筛选reads。删除重复项后的Bam文件,使用来自Sentieon DNASeq的Haplotyper [[61](#_bookmark61)]( with options –emit\_conf=20, –call\_conf=20)进行变体识别。来自青菜、白花甘蓝、合成和天然油菜组的变体使用BCF tools v 1 . 3 . 1合并 [[63](#_bookmark63)].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](#_bookmark51)]和bcftools进行单倍型感知结果的识别，对于样本中深度(深度< = 30)和质量(Q < 20)较低的基因座，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](#_bookmark64)].由ModelFinder确定的最佳模型[[65](#_bookmark65)]并为超快速引导指定1000个重复。这些树由在线工具交互式生命之树(iTOL) v3展示[[66](#_bookmark66)].GCTA v1.25对油菜材料进行了主成分分析和Fst估计[[67](#_bookmark67)].群体结构由实现了基于模型的聚类方法的STRUCTURE v2.3.4推断[[68](#_bookmark68)]。图是用R包ggplot2生成的.

**核苷酸多样性**

在3个遗传集群间的核苷酸多样性(π)和群体固定统计(Fst)用vcftools v 0 . 1 . 13计算,使用500-bp滑动窗口和100-bp步长。基于nap、cam和pol之间的核苷酸多样性比例来计算多样性的减少(ROD)值 [[4](#_bookmark13)]，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>。

附加文件1。油菜细胞器全基因组序列。

附加文件2。补充数据。

附加文件3。供应餐桌。

**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。青菜的原始序列

和白花甘蓝可在NCBI数据库的bio project accentation prjna 312457下找到。获得的用于细胞器基因组reads提取的六种芸苔属植物的已发表的cp和mt基因组列于补充数据表S[1](#_bookmark9)。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**

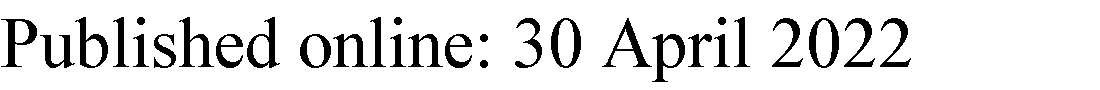
作者声明没有利益冲突。

**作者详细信息**

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



**参考**

1. 艾伦德·CJ，GJ国王。利用叶绿体和核分子标记研究油菜两倍体物种的起源。BMC Plant Biol。2010;10:54.
2. 油菜基因组分析，特别是油菜的实验形成和独特的施肥方式。Jpn J Bot。1935;7:389–452.
3. Chalhoub B，Denoeud F，Liu S，Parkin IA，Tang H，Wang X，等.植物遗传学.新石器时代后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排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 .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。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.模塑厂。2019;12:30–43.
16. 宋，刘DX，谢，杨Z，郭L，刘k，等。BnPIR: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。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 .油菜含油量的遗传分析.幼发拉底卡。2009;173:17–24.

1. 华伟，李，詹GM，刘杰，李杰，王XF，等.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.
2. 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。模塑厂。2019;12:582–96.
3. 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.
4. 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.
5. 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.
6. myszczynski K，Górski P，lipiko M，Sawicki J .对裸腹溞(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.
7. 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。
8. 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.
9. McManus HA，FuíkováK，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.

1. 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.
2. 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。
3. 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.
4. 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:对农业上重要的油菜的遗传多样性和系统发育的见解.BMC基因组公司。2020;21:480.
5. 不同有丝分裂类型在油菜线粒体基因组中共存。2011;6:e17662。
6. 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油菜异源四倍体的复杂起源和多样性。Nat Commun。2019;10:2878.
7. 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.
8. 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.
9. 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.
10. Schmutzer T，Samans B，Dyrszka E，Ulpinnis C，Weise S，Stengel D，等.来自模型异源多倍体植物油菜的物种全基因组序列和核苷酸多态性.Sci数据。2015;2:150072.
11. 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.
12. 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。
13. 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。
14. 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油菜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.

+

1. 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.
2. 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.
3. 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油菜物种之间的关系，这些物种构成了u. Front Plant Sci的三角形。2017;8:111.
4. 薛，王勇，陈明，董s，邵，刘勇. u型三角形的母体遗传与油菜植物线粒体基因组的进化过程.前线工厂科学。2020;11:805.
5. 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油菜分子标记辅助选择育种研究.植物细胞研究2017；36:901–9.
6. 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，Ostersetzer-毕然O，等. 11种芸苔素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。

1. 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.
2. 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.
3. 使用burrows- wheeler变换实现快速、准确的短reads校准。生物信息学。2009;25:1754–60.
4. 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。

1. Danecek P，McCarthy SA .BCF tols/csq:单倍型感知变异序列。生物信息学。2017;33:2037–9.
2. 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.
3. Kalyaanamoorthy S，Minh BQ，Wong T，von Haeseler A，Jermiin LS .Mod- elFinder:用于精确系统发育估计的快速模型选择。Nat方法。2017;14:587–9.
4. 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。
5. 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.
6. 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对管辖权主张保持中立。



*准备好提交您的研究了吗？选择BMC并从中受益:*

* 快速、方便的在线提交
* 由您所在领域经验丰富的研究人员进行全面的同行评审
* 验收后快速公布
* 支持研究数据，包括大型和复杂的数据类型
* 黄金开放存取，促进更广泛的合作和增加引用
* 您的研究的最大可见性:每年超过1亿的网站浏览量

**在BMC，研究总是在进行中。**

了解更多biomedcentral.com/submissions