肖等. BMC基因组学(2021) 22:262

https://doi.org/10.1186/s12864-021-07544-y

研究文章开放存取

[](http://crossmark.crossref.org/dialog/?doi=10.1186/s12864-021-07544-y%26domain=pdf)甘薯栽培种与野生种叶绿体基因组的比较分析

石卓晓1、潘旭2、易通灯1、席彬代1、陆冠昭1、贝蒂娜·海德3、安章1、支林洲1和清河草1\*

摘要

背景:甘薯Ipomoea batatas [L.] Lam.是一种重要的粮食作物。然而，由于其基因组庞大，遗传背景复杂，很难准确确定该物种核基因组的遗传信息。这一缺陷限制了对甘薯起源、进化、遗传多样性等相关研究的开展。

结果:对107个甘薯栽培种的叶绿体基因组进行了测序、组装和注释。将得到的叶绿体基因组与已发表的野生甘薯种的叶绿体基因组进行了比较分析。甘薯叶绿体基因组具有高度的相似性和一定的特异性。系统发育分析可以清楚地区分野生种和栽培种。 甘薯近缘野生种Ipomoea trifida（2x）和甘薯近缘野生种Ipomoea tabascana（4x）与栽培品种的亲缘关系最近，ycf1的不同单倍型可用于区分栽培品种与其野生近缘种。使用叶绿体基因组中的变异分析遗传结构。与传统的核标记相比，基于叶绿体基因组InDels设计的叶绿体标记具有明显优势。

结论:对107个甘薯栽培种和几个野生种的叶绿体基因组进行比较分析，有助于分析甘薯叶绿体DNA标记的进化、遗传结构和发育。

关键词:甘薯，番薯属，叶绿体基因组，比较分析，遗传结构

# 背景

甘薯Ipomoea batatas [L.] Lam.是全球重要的粮食作物，被广泛用作工业和生物能源[[1](#_bookmark14)].鉴于其相对较高的产量和较强的适应性，本种在发展中国家的粮食安全中发挥了重要作用[[1](#_bookmark14),[2](#_bookmark14)].甘薯属于旋花科甘薯属，是旋花科唯一的

六倍体(2n = 6x = 90)种[[3](#_bookmark14)].该物种基因组高度杂合，基因组大小已达到1.5 Gb，缺乏高质量和完整的参考基因组序列[[4](#_bookmark14)–[7](#_bookmark14)].迄今为止，甘薯的起源和进化仍不清楚[[8](#_bookmark14)].现代甘薯被推测为是第二次全基因组复制后四倍体祖先和二倍体祖先的初代杂交，[[7](#_bookmark14)].甘薯近缘野生种Ipomoea trifida是最可能的二倍体祖先，这一观点得到了其全基因组测序的支持

[[6](#_bookmark14)]，然而，四倍体祖先仍然是未知的。

叶绿体是植物的关键细胞器。除了在光合作用中众所周知的功能外，叶绿体还参与重要的生物过程，如植物免疫和作物品质[[9](#_bookmark14),[10](#_bookmark14)].叶绿体的遗传转化已成为基因工程研究的热点[11](#_bookmark14)].叶绿体基因组是一个封闭的环状DNA，在细胞中以多拷贝的形式存在。高等植物的叶绿体基因组具有高度保守的四方环状结构，大小在115-165 kb之间。两个反向重复(IR)序列将整个环状叶绿体基因组分成一个大的单拷贝(LSC)和一个小的单拷贝(SSC) [[12](#_bookmark14),[13](#_bookmark14)].

叶绿体基因组含有重要的遗传信息。叶绿体基因组的编码区和非编码区在分子进化速度上有显著的差异，这些差异适用于不同类别的系统研究[[14](#_bookmark14)].此外，叶绿体DNA (cpDNA)的核苷酸取代率适中，叶绿体基因组的大小不是很大，便于测序。不同物种的叶绿体基因组具有良好的共线性，使得叶绿体基因组的数量更容易组装。特别是对于具有复杂核基因组的物种，如甘薯来说,叶绿体基因组的这些优势更为显著，。因此，基于叶绿体基因组的质体组学方法近年来发展迅速[[15](#_bookmark14)–[18](#_bookmark14)].

中国是全球最大的甘薯生产国，年产量为5324.57吨，占世界总产量的57.91 %[[19](#_bookmark14)].上个世纪，中国广泛种植的甘薯栽培种主要是日本的‘Okinawa 100’和美国的‘Nancy Hall’及其衍生品种或后代，如Xushu 18,这是中国种植最广泛的甘薯栽培种，已经成为许多中国的流行品种的亲本

[[20](#_bookmark14),[21](#_bookmark14)].中国甘薯的遗传背景相对狭窄[20](#_bookmark14)]，但随着材料数量的不断增加，有必要对甘薯栽培种进行分子鉴定和多样性分析。使用分子标记，如简单序列重复和扩增片段长度多态性分析了甘薯的遗传多样性[[21](#_bookmark14),[22](#_bookmark14)].考虑到甘薯的多倍体特性，这些标记的特异性并不理想。基于序列的单核苷酸多态性(SNP)和特定长度的扩增片段可以提高标记物的密度[[20](#_bookmark14),[23](#_bookmark14)].然而，由于缺乏高质量的参考基因组，突变识别的可靠性值得商榷。在准确获得核基因组的细节之前，利用叶绿体基因组分析甘薯的遗传多样性是一个很好的选择。

本研究对107个甘薯栽培种的叶绿体基因组进行了测序和组装。结合已发表的11个野生种的叶绿体基因组，进行了比较基因组、系统进化和遗传结构分析。基于叶绿体基因组中的插入-缺失(InDel)变体设计了用户友好的分子标记。该结果为甘薯的类组学、遗传进化和精确分子鉴定研究奠定了基础。

# 结果

107个甘薯栽培种全基因组重测序及叶绿体基因组组装

在全球范围内共获得107个甘薯栽培种，其中92个样品来自中国各地(图[1](#_bookmark0)桌子，[S1](#_bookmark8)).利用Nova- Seq 6000平台对107个甘薯栽培种的全基因组进行重测序，获得原始数据2064.03 Gb。经过过滤后，得到以下数据:干净数据2056.36 Gb，平均每个样本19.22 Gb,平均测序深度，

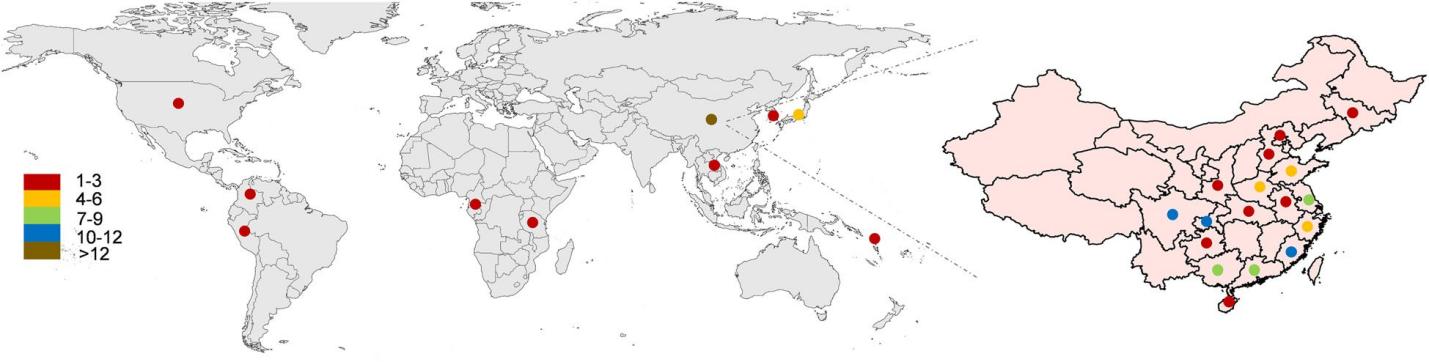


图1甘薯样品的地理分布和数量。不同的颜色表示从该区域采集的样本数量。左侧是全球样本，右侧是中国样本。这些地图是使用以下R包绘制的:“maptools 1.0–2”，“map3.3.0”，" ggmap 3.0.0" 和" mapdata 2.3.0"

超过12倍；Q20，95.68–97.85%；Q30，89.22–

93.73%;GC含量36.49 ~ 39.41%，平均GC含量37.54%。(表[S1](#_bookmark8)).

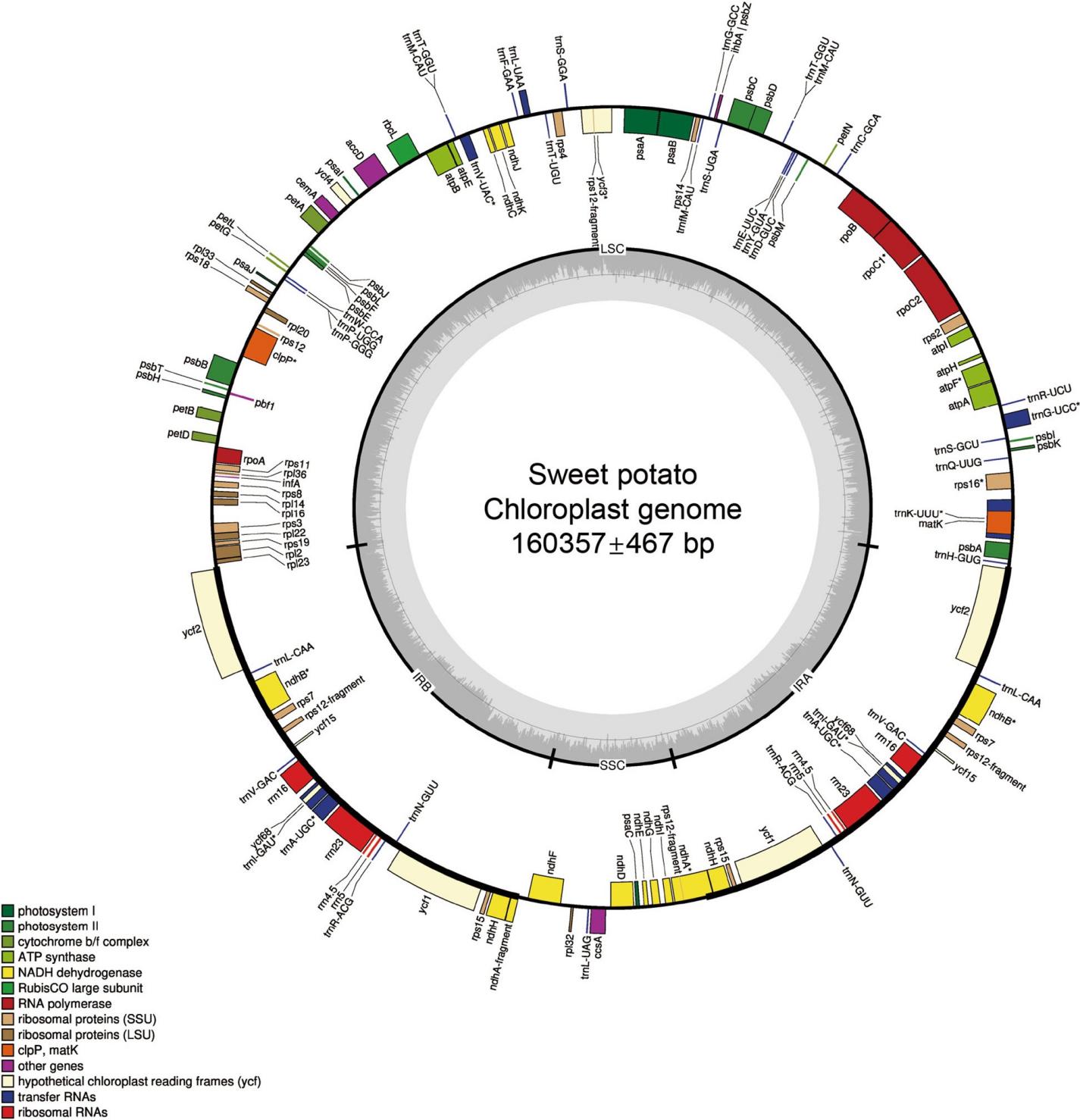
选择 Xushu18的叶绿体基因组作为参考基因组，筛选比对reads进行组装。Contigs N50和N90的平均长度分别为77，921 bp和11，343 bp，scaffolds N50和N90的平均长度

分别达到85,023 bp和16,431 bp，不过平均gaps数为1.88(表[S1](#_bookmark8)).选择长scaffolds拼接成环状DNA。

甘薯叶绿体基因组结构

甘薯叶绿体基因组具有大多数高等植物所特有的四方结构。这些叶绿体基因组的长度在156，

图2 107份甘薯叶绿体基因组图谱。画在圆圈内的基因是顺时针转录的，圆圈外的基因是逆时针转录的。内环中较深的灰色表示GC含量。LSC，大单副本；SSC，短单份；IR，反向重复序列。叶绿体基因组的长度标记在中心

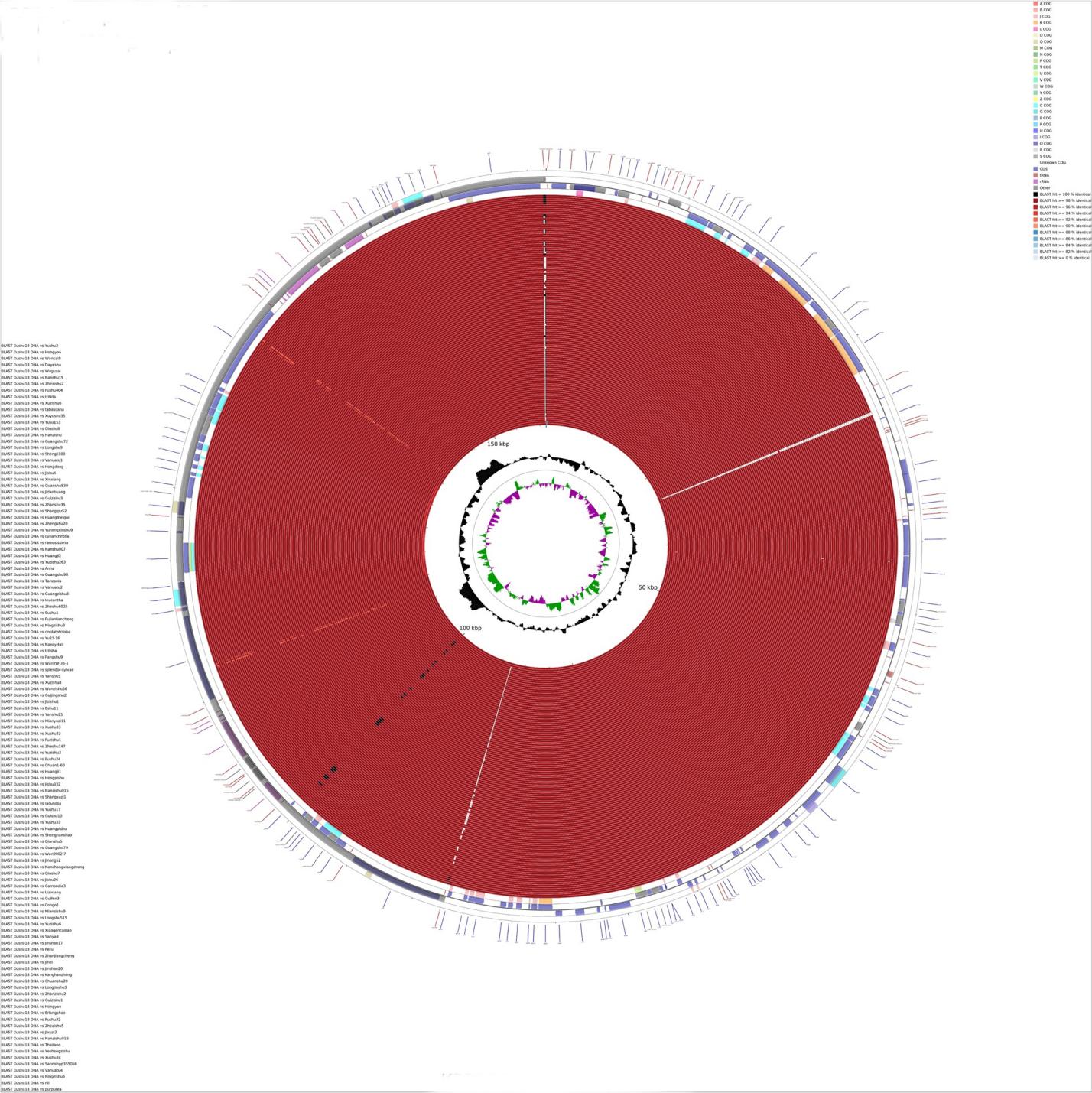


888 bp和161，302 bp，with Med = 87,754 bp and x = 87,791 ± 157 bp (Fig. 2). The LSC (length = 87,589–88, 298 bp, Med = 87,754 bp and x = 87,791 ± 157 bp) and SSC (length = 12,047–12,143 bp, Med =12,065 bp and x = 12,068 ± 21 bp) were separated by two IRs (length = 26,923–30,675 bp, Med = 30,226 bp and x= 30,220 ± 347 bp). 这些甘薯和参考品种 Xushu18的叶绿体基因组表现出良好的同源性，

这证明了甘薯Ipomoea batatas [L.] Lam.叶绿体基因组的保守性(图[S1](#_bookmark6)).

甘薯完整的叶绿体基因组包含80个编码蛋白的基因，其中8对位于2个IR区。共发现37个tRNAs，其中7个在每个IR区域中发现。此外，还观察到22个包含内含子的orf，由14个蛋白编码基因(ndhA，ndhB，

图3 107份甘薯与11份野生种叶绿体基因组比较分析。特征环(从最外环开始)的内容如下:环1:正向链编码序列的COG功能类别；环2:正向链序列特征；环3:反向链序列特征；环4:反向链编码序列的COG功能类别。接下来的118个环显示了通过BLAST检测到的序列相似性区域，BLAST是通过参考基因组和118个叶绿体基因组之间的DNA序列进行比较的。最里面的两个环分别显示GC偏斜和GC含量



rps16、rpoC1、ycf3、clpP、petB、petD、rpl16、atpF及两拷贝的ndhB和rps)和八个trnA(trnK-UUU、trnG-UCC、trnL-UAA、trnV-UAC及两拷贝trnI- GAU和trnA-UGC)(图。[2](#_bookmark1)).

**甘薯叶绿体基因组的比较分析11种甘薯的叶绿体基因组**从NCBI下载11个甘薯属物种用于与107种甘薯栽培品种进行叶绿体基因组比较分析(图[3](#_bookmark2)).在不同品种甚至不同种的甘薯属植物中，大部分核苷酸序列的相似性都高于98%。然而，也存在一些例外情况，即在LSC约115–136 kb处，与其他甘薯属相比，Ipomoea nil and Ipomoea purpurea的相似性低于98%，高于94%。该区域包含整个SSC、位于两个IRs中的rps15和ndhH。这两个野生物种属于

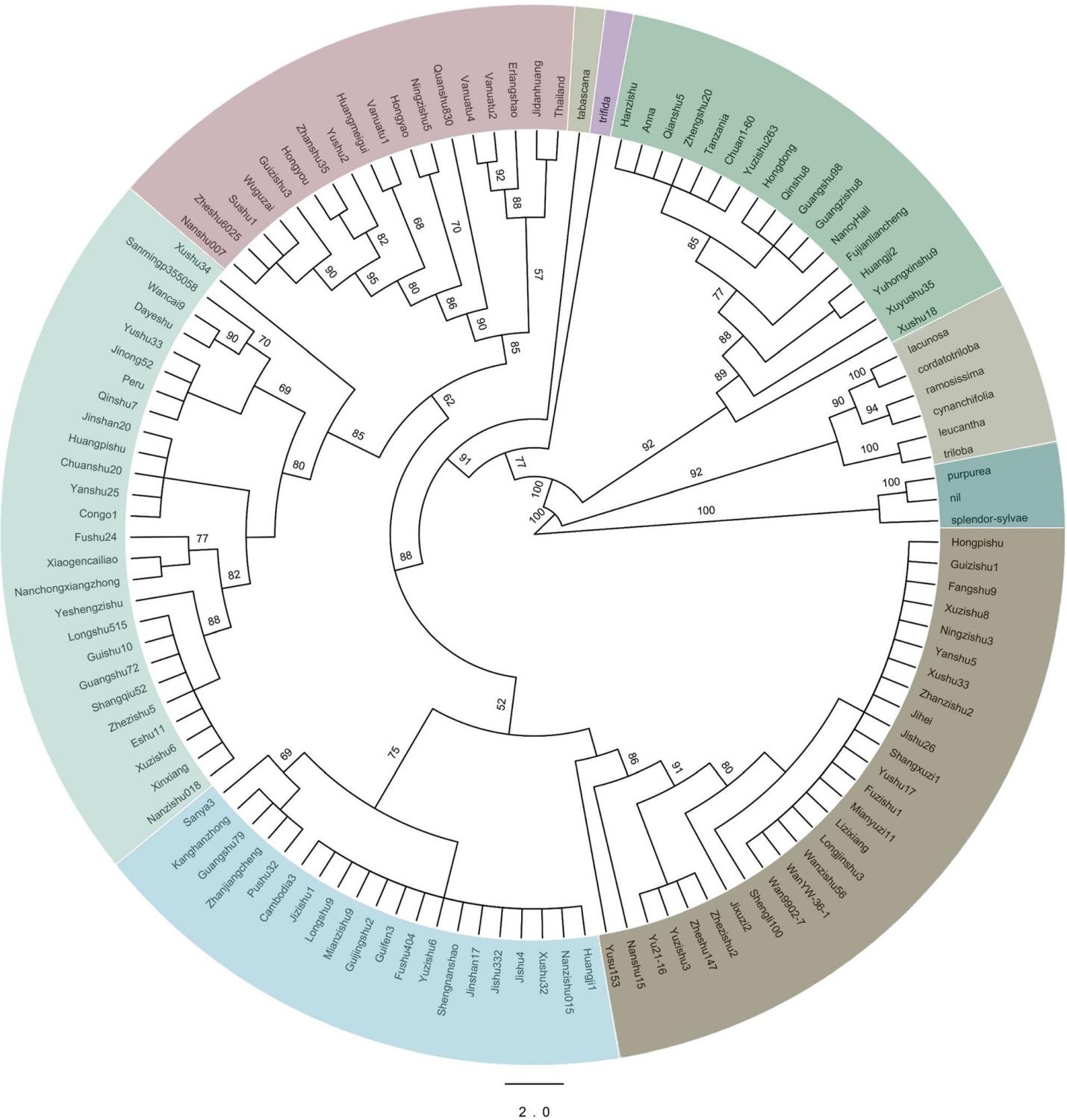


图4基于叶绿体单拷贝基因的107个甘薯栽培种和11个野生种系统发育树。不同的分支颜色不同，分支I、II、IV和V为野生种，其他为栽培种

茑萝属Quamoclit(旋花科旋花亚科番薯族茑萝属),被认为与甘薯栽培种(旋花科旋花亚科番薯族番薯属番薯)有相对远缘关系。叶绿体基因组这一部分的差异也可以支持这一假设。此外，ycf1基因的两个拷贝中存在明显的低相似性位点。ycf1基因编码一种具有未知功能的蛋白质，该蛋白质可能作为复合物的组分参与蛋白质转运[[28](#_bookmark14)].

考虑到ycf1基因的多态性，从部分栽培种和11个野生种中筛选出ycf1，并对氨基酸(AA)序列进行比对(图[S2](#_bookmark7)).结果显示，该基因上存在一个以SEKKSETD序列为单位的重复区域，范围为1755 AA至1810 AA。在栽培种中发现了四个重复序列，在甘薯的野生种中发现了五个或六个重复序列。此外，在I.nil and I.purpurea中发现了7个重复序列，在I.purpurea中发现了2个突变，因此ycf1在栽培种、野生近缘种和相对远缘的野生种中有其特定的单倍型。

甘薯的系统发育分析

基于叶绿体基因组中注释的单拷贝基因，用最大似然法(ML)对11个野生种和107个甘薯栽培种进行了系统发育分析(图。[4](#_bookmark3)).这些样本被明确分为九个分支。第一分支，包括

三种野生物种(I.purpurea, I.nil and I.splendor-sylvae)离栽培品种最远。第二分支由6种野生种(I.triloba, I.lencantha, I.cynanchifolia,

I.ramosissima, I.cordatotriloba and I.lacunosa)组成，均属于甘薯部分。甘薯部分组成了与甘薯栽培种最近的关系。其他两种野生物种(I.trifida and I.tabascana)被分为两个独立的分支:分支IV和V，它们与栽培种的关系最为密切。二倍体物种I.trifida长期以来被认为是甘薯的祖先之一[[6](#_bookmark14)].该系统发育分析可能支持这一观点。

Xushu18曾是中国种植最广泛的品种，其亲本之一‘NancyHall’被归为三号枝。该分枝中的大多数品种与 Xushu18或其亲本有相对来说的亲缘关系。与其他分支相比，第六分支表现出较大的多样性，尽管该分支中仅有18个品种，但这些品种被划分为9个亚分支。该分支的品种来自多个国家，包括日本、泰国、瓦努阿图，还包括一些中国古代的地方品种。分枝VII和IX是分别包含26和27个甘薯栽培种的两个最大分枝。第VII分支主要是橘黄果肉的

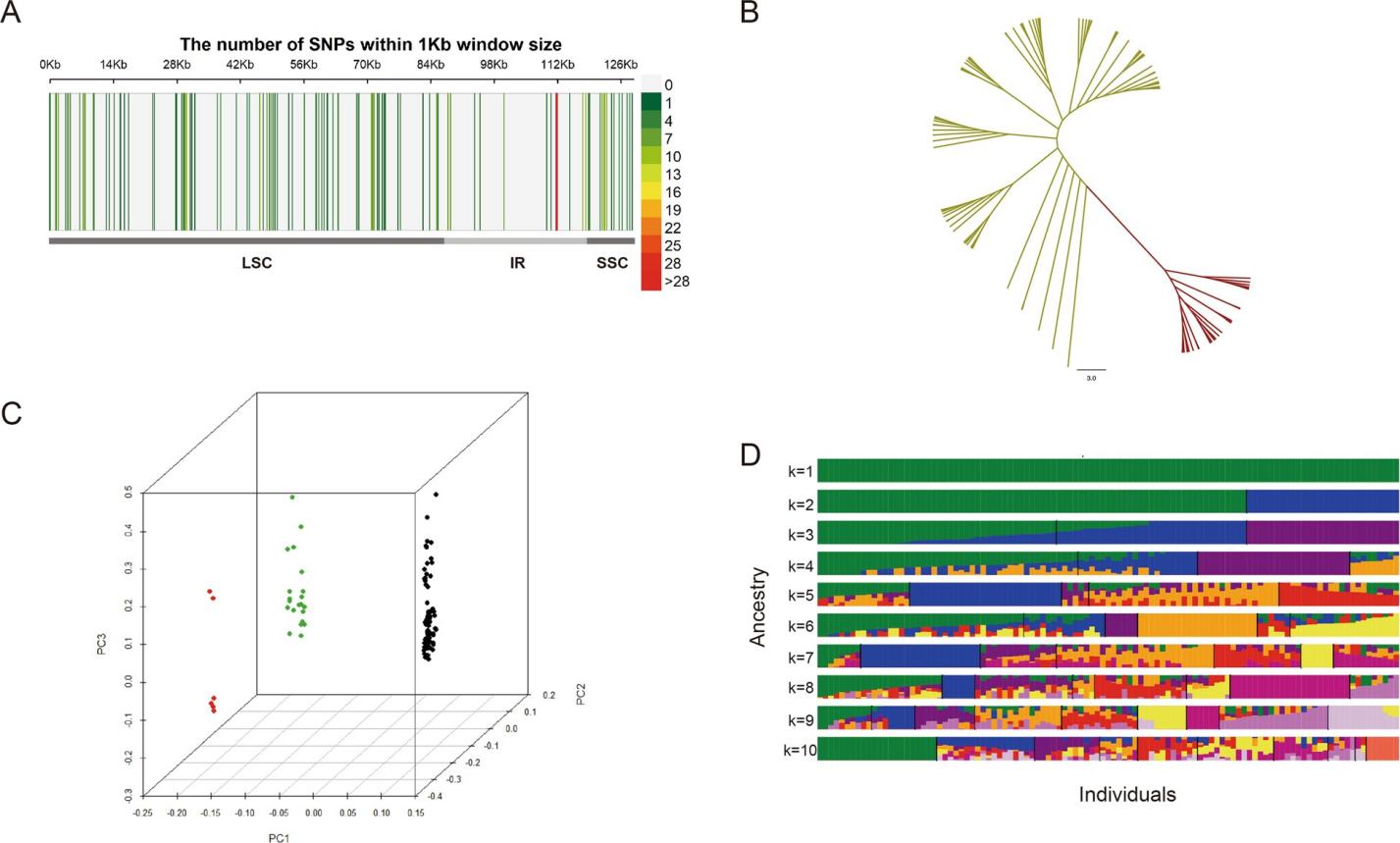


图5基于叶绿体基因组变异的107个甘薯栽培种遗传结构分析。叶绿体基因组上SNP的密度分布(一):系统发育树(b)；主成分分析(c)；K范围为1至10 (d)的群体结构分析

甘薯，占26个栽培种中的18个。相比之下，第九分支的大部分品种是紫色甘薯，占27个栽培种中的15个。第八分支的标本主要来源于中国南方沿海地区，其中广东省3个，福建省3个，广西省2个，海南省2个(图。[4](#_bookmark3)桌子，[S1](#_bookmark8)).

构建了基于ycf1的系统发育树(图[S3](#_bookmark9)).基于叶绿体基因组的系统发育树与品种分类没有明显的相关性。但甘薯的野生近缘种仍清晰地形成一个分支，表明该基因代表了甘薯属的不同类群。

因此，尽管甘薯栽培种被分成不同的分支，但没有观察到地理、果肉颜色等性状与分支之间的显著关系。这种结果可能是由于中国不同地区品种之间广泛存在的相互引进和杂交所造成的。

突变识别和遗传结构分析

以 Xushu18叶绿体基因组为对照，检测甘薯cpDNA变异体。共筛选出229个突变位点，其中118个SNPs和111个InDels(图。[5](#_bookmark4)a)。在这些变体中，129个变体位于基因的上游和下游(66个SNPs和63个indel)，3个变体位于基因间区(均为SNPs)，31个变体位于ncRNA或内含子中(25个SNPs和6个indel)。外显子中共有66个变体(54个SNPs和12个InDels)，其中25个是非同义突变(表[S2](#_bookmark10)).基因ycf1携带多达31个突变位点,

与比较基因组学分析一致(图[3](#_bookmark2)).

提取的SNPs用于遗传结构分析。进行了系统发育树的构建、主成分分析和群体结构分析(图[5](#_bookmark4)).系统发生树显示两个主要类群明显聚集(图[5](#_bookmark4)b)。当使用3个主成分时，107个品种可分为3组，大部分样本分为2个最大组，最小组只有6个样本(图[5](#_bookmark4)c)。用1 ~ 10的K值分析群体结构，K = 2时群体明显分离(图[5](#_bookmark4)d)。K = 2时，交叉验证(CV)误差也最低(图。[S4](#_bookmark11)).总之，这些发现表明最好将甘薯群体分为两组。

叶绿体DNA标记的研究进展

基于叶绿体基因组上检测到的InDels，选择碱基数不低于3的位点变异设计cpDNA标记。设计了20对扩增引物。这些引物的长度在20 bp至26 bp之间，Tm值在57.47°C至60.42°C之间。正向引物和反向引物之间的最大Tm差异为1.95°C。产物的长度大多在130 bp至195 bp之间，除了Ibcp-15的产物达到300 bp(表[S3](#_bookmark12)).在cpDNA标记中，13个来自LSC，4个来自IR区域，3个来自SSC。

为验证cpDNA标记的有效性，随机抽取8个甘薯栽培种，提取其DNA作为模板。PCR结束后进行毛细管电泳。与核DNAmarkers相比,

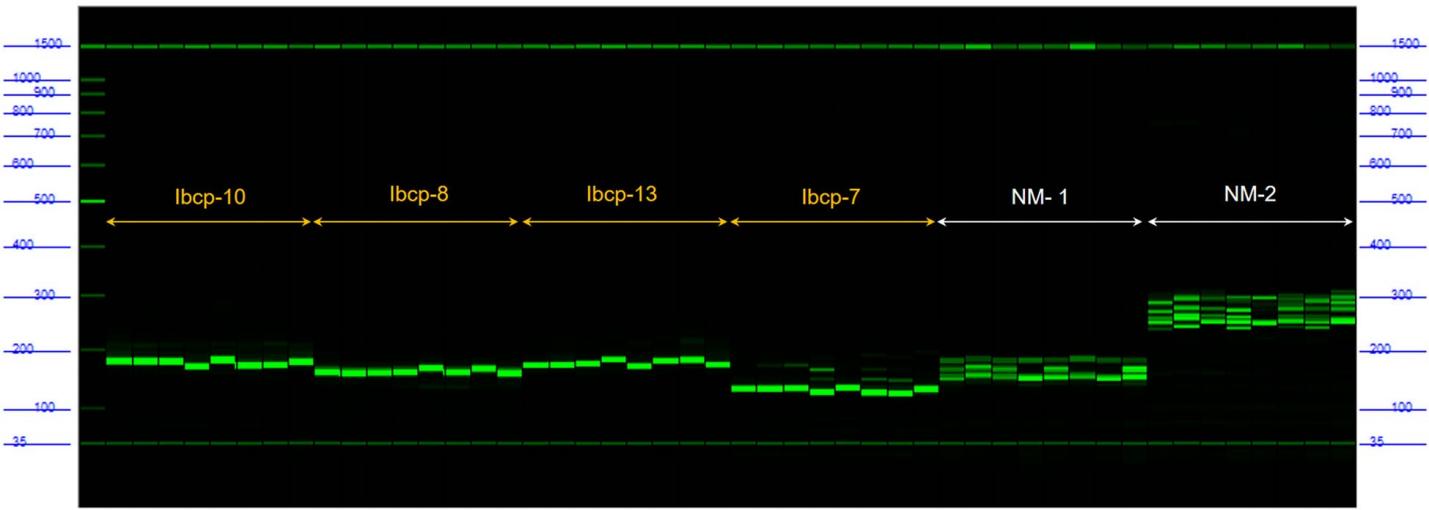


图6采用cpDNA标记的毛细管电泳。左侧4组为cpDNA标记扩增产物(Ibcp-10、Ibcp-8、Ibcp-13、Ibcp-7)，右侧2组为核标记扩增产物(NM-1、NM-2)

cpDNA标记条带单一，清晰，可读性更强(图[6](#_bookmark5)).而且这些标记在样本间表现出良好的多态性。因此，这些优秀的cpDNA标记可为甘薯遗传多样性分析或甘薯栽培种指纹图谱的构建提供有力的工具。

# 讨论

本研究对107个甘薯栽培种进行了测序。鉴于甘薯的基因组大小，对大量样本进行重新测序是一项艰巨的任务。目前，可用的甘薯属参考基因组来自二倍体野生近缘种:I.trifida [4, 6, 29], I.triloba [6], and I.nil [5]，以及来自栽培品种Taizhong 6 的单倍型基因组[[7](#_bookmark14)].由于缺乏高质量的参考基因组，重测序自然群体的应用受到限制。由中国、日本和韩国共同努力推广的甘薯栽培种 Xushu18的重新测序将在不久的将来完成[[30](#_bookmark14),[31](#_bookmark14)].使用甘薯的重测序群体进行全基因组关联分析或泛基因组分析是一个机会。

cpDNA参与了植物体内一些重要的生理过程。cpDNA所携带的信息有助于更全面地了解植物的遗传和进化。本研究从大量的DNA序列中提取出cpDNA片段，完成同源性组装和注释，获得107个优质叶绿体基因组。鉴于甘薯核基因组的复杂性和获得核基因组信息的困难性，全面了解叶绿体基因组则是一个有价值的补充。

甘薯叶绿体基因组均为典型的四方环状结构，在属内物种间高度保守，尤其是相似度大于98%的核酸序列。尽管如此，在LSC 115 ~ 136 kb的区域内，仍可明显区分出I.nil和I . purium(图[3](#_bookmark2)).这种特征可能是系统发育分析中I.nil 和 I.purpurea被划分为单个分支的原因(图[4](#_bookmark3)).

无论是比较基因组学分析还是突变识别，都已经证明了ycf 1携带大量的突变(图1和图2)。[3](#_bookmark2)和[5](#_bookmark4)a，桌子[S2](#_bookmark10)).前期研究表明，ycf1在不同物种间进化速度较快。因此，该基因被用作cpDNA条形码来鉴别不同的物种[[32](#_bookmark14),[33](#_bookmark14)].在本研究中，ycf1不仅在不同物种之间有很大差异，在不同甘薯栽培种之间也有很大差异(图[3](#_bookmark2)).鉴于

在ycf1中有相当大的变异，ycf1在一些物种中已成为假基因，并且ycf1的丢失在植物中是常见的，因此ycf1是否是不可缺少的是有争议的[[34](#_bookmark14)–[36](#_bookmark14)].本研究结果表明，甘薯ycf1具有完整的开放阅读框，能够正常编码蛋白质。此外，其AA序列还可用于甘薯与其野生近缘种的鉴别。[S2](#_bookmark7)和[S3](#_bookmark9)).

根据叶绿体基因组的单拷贝基因，将11个野生种和107个甘薯栽培种分成9个分支(图[4](#_bookmark3)).野生种与栽培种明显分离，其中4个分支由野生种组成，5个分支由栽培种组成。在野生种中，I.tabascana and I.trifida被分成两个独立的分支，这两个野生种与栽培种的关系比任何其他野生种都密切。甘薯栽培种的起源长期以来一直不明确，但普遍认为I.trifida是甘薯的祖先之一，这已被细胞学标记和染色体标记所证实[[8](#_bookmark14),[37](#_bookmark14)].本研究还利用叶绿体比较基因组证实了I.trifida是一种与栽培种关系相对密切的野生种(图[4](#_bookmark3)).然而，四倍体是否参与进化，以及哪些四倍体参与进化，仍然存在争议。虽然本研究不能完全解释这一问题，但四倍体I.tabascana在栽培种附近形成分支的结果可能为甘薯的进化提供一些有用的信息。

以前使用染色体分子标记研究过甘薯栽培种的遗传多样性和群体结构。利用38个SSR位点和62，363个SNPs位点分别对中国甘薯材料的遗传多样性和群体结构进行了分析，并将这些甘薯栽培种分为3组[[20](#_bookmark14),[38](#_bookmark14)]，与本研究PCA结果一致(图[5](#_bookmark4)c)。然而，由叶绿体基因组上的单拷贝基因和突变位点构建的树之间存在差异(图1和图2)。[4](#_bookmark3) 和[5](#_bookmark4)b)可能是由cp基因组上的编码区和非编码区的进化速率的差异引起的。229个变异位点中只有25个非同义变异也支持这一观点(表[S2](#_bookmark10)).此外，部分栽培种在叶绿体基因组水平上没有得到很好的区分(图[4](#_bookmark3))，表明甘薯叶绿体基因组具有种内保守性。

DNA分子标记是遗传多样性评价和分子指纹构建的基础。甘薯是六倍体，因此多个拷贝和杂合位点导致普通核标记的可用性非常差。基因组的放大位置

并不清晰和特异，以至于一对引物甚至可以在电泳过程中产生多达20条条带，导致电泳结果可读性很差。本研究根据InDel 识别设计了20个cpDNA标记。通过防止引物与核DNA中的任何地方结合，确保了引物的特异性。进行毛细管电泳。与核标记相比，甘薯cpDNA标记产物具有单一、清晰、特异性好、可读性强的特点(图[6](#_bookmark5)).cpDNA标记是分析甘薯遗传多样性或构建甘薯栽培种指纹图谱的有力工具。除InDels外，在cpDNA中也发现了丰富的SNPs(表[S2](#_bookmark10))，其可被设计为与甘薯cpDNA标记互补的dCAP标记。这些结果将有助于提高我国甘薯栽培种的同质性。

# 结论

本研究对107个甘薯栽培种的叶绿体基因组进行了测序、组装和注释。对107个甘薯栽培种和野生种的叶绿体基因组进行了比较分析。甘薯栽培种及其野生种在叶绿体基因组中保持高度相似性。叶绿体基因组可以清晰地区分栽培种和野生种。I.trifida and

I.tabascana 与栽培变种关系最为密切，可能参与了甘薯的进化。甘薯栽培种明显属于几个群体，但与地理来源或果肉颜色没有显著关系。基于叶绿体基因组变异设计的cpDNA标记与传统的核标记相比具有显著优势。该设计标记可用于甘薯栽培种的遗传多样性分析和分子鉴定。

# 方法

植物材料和再测序

在全球范围内共获得107个甘薯栽培种，并保存在徐州国家甘薯基因库。所有甘薯种质资源均为公共品种或地方品种。在这些种质资源中，92种来自中国，其余来自美国、柬埔寨、刚果、日本、秘鲁、韩国、坦桑尼亚、泰国和瓦努阿图(表[S1](#_bookmark8)).对这些品种的鲜叶取样，并使用液氮研磨成粉末。总DNA采用CTAB法提取[[39](#_bookmark15)].使用NanoPhotometer分光光度计(uniplen，CA，USA)检查DNA纯度。DNA浓度

使用Qubit 2.0荧光计(Life Technologies，CA，USA)中的Qubit DNA分析试剂盒测量。每份样本中总计700 ng DNA被用作DNA样本制备的输入材料。使用NEB Next Ultra DNA文库制备试剂盒(Illumina，NEB，USA)按照制造商的建议生成测序文库，并将索引代码添加到每个样本的属性序列中。使用AMPureXP系统(美国贝弗利贝克曼库尔特)纯化DNA。DNA片段3’端腺苷酸化后，将携带发夹环结构的NEB Next Adaptor连接起来，为杂交做准备。然后通过电泳筛选出具有特定长度的DNA片段。在PCR前，使用3 μL USER酶(NEB，USA)选择大小，在37°C下将连接的DNA加热15 min，然后在95°C下加热5 min。然后使用Phusion高保真DNA聚合酶、通用PCR引物和Index (X)引物进行PCR。最后，纯化PCR产物(AMPure XP系统)，并在Agilent bio analyzer 2100系统上评估文库质量。合格的文库用于在NovaSeq 6000平台上测序。插入片段应为350 bp，并产生150 bp的双端测序。

甘薯叶绿体基因组的组装和注释

使用fastp对测序数据进行质量控制[[40](#_bookmark15)].HISAT2[[41](#_bookmark15)]用于将筛选的读数与参考叶绿体基因组对齐[[42](#_bookmark15)].用spades将比对的reads拼接成scaffolds[[43](#_bookmark15)].scaffolds和参考叶绿体基因组之间的共线性分析是通过MUMmer4.0进行的[[44](#_bookmark15)].然后选择优质的scaffolds组装成环状的DNA分子。同源性注释在线生成[[45](#_bookmark15)].GeSeq模块[[46](#_bookmark15)]用于注释环状DNA，并手动优化结果。另一个模块OGDRAW [[47](#_bookmark15)]被用来绘制圈图。

甘薯比较基因组学分析

11种野生甘薯(Ipomoea trifida, Ipomoea tabascana, Ipomoea triloba, Ipomoea cordatotriloba, Ipomoea cynanchifolia,Ipomoea splendor-Sylvae, Ipomoea ramosissima, Ipomoea leucantha, Ipomoea lacunosa, Ipomoea nil, Ipomoea purpurea)叶绿体基因组序列[[48](#_bookmark15)]和栽培品种 Xushu18 [[49](#_bookmark15)]从NCBI下载[[50](#_bookmark15)].将11个野生种和107个栽培种的基因库文件导入CGView比较工具[[51](#_bookmark15)]，脚本“build\_ blast\_atlas.sh”用于自动创建用于核苷酸(blastn)比较的图谱。ycf1的AA序列通过geneius Basic 4 . 8 . 5进行比对[[52](#_bookmark15)].

系统发育树的构建

为了识别基因家族，OrthoFinder (v 2.3.14)流程[[53](#_bookmark15)]按顺序应用于具有全对全BLASTP (E值≤1e-5)、相互最佳命中、通过同源性和辅助连接的对的十个基因组，通过OrthoFinder对E值和聚类进行归一化。最后，基因被分为直系同源物、旁系同源物和单拷贝直系同源物(每个物种中只有一个基因)。利用单拷贝同源基因构建系统发育树,使用Mafft对每个基因家族核苷酸序列进行比对，并使用FastTree和MrBayes分别用最大似然和贝叶斯推断(BI)构建系统发育树[[54](#_bookmark15)] [[55](#_bookmark15)].比较两种方法的分支图，认为ML法构建的进化树更适合(图[4](#_bookmark3)图和[S5](#_bookmark13)).系统发育树由Figtree可视化和修改[[56](#_bookmark15)].

突变识别

Bowtie2 [[57](#_bookmark15)]用于将clean data与参考叶绿体基因组比对。使用SAMtools和BCFtools进行变体识别[[58](#_bookmark15),[59](#_bookmark15)].然后使用VCFtools过滤SNPs和InDels[60](#_bookmark15)],缺失率低于50%，次要等位基因计数高于3，次要等位基因频率高于0.05。变体的作用由ANNOVAR评估[[61](#_bookmark15)].

系统发育树、群体结构分析和基于变异体的PCA

使用筛选的SNP分析群体结构。CV误差采用ADMIXTURE[[62](#_bookmark15)]默认参数从K = 1到K = 10评估。通过R包(barplot)进行可视化。PCA由Plink进行[[63](#_bookmark15)]，并用R软件包((scatterplot3d)绘制三维图形。选择FastTree，用ML法构建系统发育树。Figtree用于可视化。

cpDNA标记物的开发与验证

选择碱基个数相差3个以上的InDels设计扩增引物。Primer3同时设计了多对引物[[64](#_bookmark15),[65](#_bookmark15)].参数设置如下:产物长度，小于200 bp;Tm值，58-64℃；正向引物和反向引物之间的Tm值差异，小于2℃；GC含量，35-65%；引物的长度分布在20 bp和26 bp之间。由生工 生物技术公司选择得分最高的引物进行核苷酸序列合成。8个甘薯栽培被随机选取

抽取DNA作为模板，使用设计的引物进行通用PCR，并设置2个核标记作为对照。PCR后在 Fragment

Analyzer 系统 (AATI, USA)上进行毛细管电泳，在计算机上读取条带。

# 补充信息

在线版本包含补充材料，可查阅https://doi。org/10.1186/s12864-021-07544-y。

附加文件1:图S1。品种与 Xushu18叶绿体基因组的同源性分析

附加文件2:图S2。部分甘薯栽培种和野生种ycf1氨基酸序列的比对。前5行是栽培品种，后9行是Batatas剖面的野生种，最后两行是Quamoclit剖面的野生种。

附加文件3:图S3。基于ycf1的107个甘薯和11个野生种系统发育树。分枝被染成红色是巴塔塔斯节的野生物种。

附加文件4:图S4。不同K值的CV误差。

附加文件5:图S5。BI法构建基于单拷贝基因的系统发育树。

附加文件6:表S1。107个甘薯栽培种的详细资料。

附加文件7:表S2。cpDNA的变异。

附加文件8:表S3。基于Indels设计的cpDNA标记

承认

我们要感谢徐涛博士的仔细审阅和建议。

作者贡献

QC和SX设计了这个实验。SX完成了数据分析并撰写了手稿。PX协助进行了本研究的构思和数据分析。YD制备了用于DNA提取和测序的样本，并设计了叶绿体DNA标记。XD和AZ负责甘薯样品的采集和保存，以及样品全基因组测序的辅助工作。LZ进行毛细管电泳。BH帮助组装叶绿体基因组并指导手稿的写作。ZZ促成了突变识别。ZZ和QC负责监督研究并修改手稿。所有作者都已阅读并认可了手稿。

提供资金

本工作得到了国家重点研发计划项目(2018YFD1000705/2018YFD1000700)、中国农业研究体系(CARS-10-B1)、徐州农业科学院研究基金项目(XM2020002)的资助。

数据和材料的可用性

用于分析的107个品种的测序数据保存在NCBI，注册号为PRJNA715261。从NCBI可获得用于分析的11种野生甘薯和 Xushu18的叶绿体基因组数据，其登录号为:I.trifida，MH173261I .三叶虫，MH173262I .亮氨酸塔，NC \_ 041208I.tabascana，NC \_ 041207I .星辉-sylvae，NC \_ 041206I.ramosissima，NC \_ 041205I.cordatotriloba，NC \_ 041204一、白薇，NC \_ 041203I .腔隙性脑梗死，MH173257I.nil，AP017304I .紫色，EU118126 Xushu18，NC\_026703。

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中国农业科学院江苏徐州甘薯研究中心/甘薯研究所，徐州221131。2兰州大学畜牧农业科技学院，兰州730020。3国际马铃薯中心，Av。La Molina 1895，La Molina，利马，秘鲁。

收到:2021年1月1日接受:2021年3月22日



参考

1. Zhang H，Zhang Q，翟h，Gao S，Yang L，Wang Z，等. IbBBX24促进茉莉酸途径，增强甘薯抗枯萎病能力.植物细胞。2020;32(4):1102–23.<https://doi.org/10.1105/tpc.19.00641>。
2. 刘淇。甘薯农艺性状的基因工程改良。品种Sci。2017;67(1):15–26.<https://doi.org/10.1270/jsbbs.16126>。
3. 张l，于勇，石涛，寇姆，孙杰，徐特，等.表达数量性状位点的全基因组分析揭示了甘薯贮藏根基因表达变异的调控结构.霍蒂奇雷斯-英格兰。2020;7(1).<https://doi.org/10.1038/s41438-020-0314-4>。
4. Hirakawa H，Okada Y，Tabuchi H，Shirasawa K，Watanabe A，Tsuruoka H，et al .

名词（noun的缩写）伊泽贝·s .野生甘薯基因组序列调查。DNA第2015号决议；22(2):171–9.<https://doi.org/10.1093/dnares/dsv002>。

1. Hoshino A，Jayakumar V，Nitasaka E，Toyoda A，Noguchi H，Itoh T，等.日本牵牛花甘薯基因组序列及分析Nat Commun。2016;7(1).<https://doi.org/10.1038/ncomms13295>。
2. Wu S，Lau KH，Cao Q，Hamilton JP，Sun H，Zhou C，等。栽培甘薯的两个二倍体野生近缘种的基因组序列揭示了遗传改良的目标。Nat Commun。2018;9(1).<https://doi.org/10.1038/s41467-018-06983-8>。
3. Yang J，Moeinzadeh M，Kuhl H，Helmuth J，Xiao P，Haas，等.单倍型解析甘薯基因组可追溯其六倍化历史.纳特植物。2017;3(9):696–703.<https://doi.org/10.1038/s41477-017-0002-z>。
4. 希尔亚克-雅科夫列夫。甘薯的起源和进化。)及其野生近缘种。植物科学。2006;171(3):424–33.<https://doi.org/10.1016/j.plantsci.2006.05.007>。
5. Medina-Puche L，Tan H，多格拉V，Wu M，Rosas-Diaz T，Wang L，等.连接质膜和叶绿体并被病原体共同选择的防御通路.手机。2020;182(5):1109–24.<https://doi.org/10.1016/j.cell.2020.07.020>。
6. Llorente B，Torres-Montilla S，莫雷利L，Florez-更纱I，Matus JT，埃斯克罗M，等。叶片叶绿体向富含类胡萝卜素质体的合成转化揭示了天然质体发育的机理基础。美国国家科学院学报2020；117(35):21796–803.<https://doi.org/10.1073/pnas.2004405117>。
7. Ruf S，Forner J，哈塞C，Kroop X，Seeger S，Schollbach L，等.可育移植拟南芥植物的高效生成.纳特植物。2019;5(3):282–9.<https://doi.org/10.1038/s41477-019-0359-2>。
8. Jansen RK，Raubeson LA，Boore JL，DePamphilis CW，Chumley TW，Haberle RC，等.获取和分析叶绿体全基因组序列的方法.方法酶法。2005;395:348–84.<https://doi.org/10.1016/S0076-6879(05)95020-9>。
9. 帕尔默·JD。叶绿体基因组的比较组织。Annu Rev Genet。1985;19(1):325–54.<https://doi.org/10.1146/annurev.ge.19.120185.001545>。
10. 克莱格MT，高特BS，学习GJ，莫顿BR。叶绿体DNA进化的速率和模式。美国国家科学院学报，1994；91(15):6795–801.<https://doi.org/10.1073/pnas.91.15.6795>。
11. Magdy M，Ou L，Yu H，Chen R，Zhou Y，Hassan H，等.泛质体方法可用于评估栽培辣椒物种的遗传变异.霍蒂奇雷斯-英格兰。2019;6(1).<https://doi.org/10.1038/s41438-019-0191-x>。
12. Carbonell-Caballero J，Alonso R，ibanez V，特洛尔J，Talon M，Dopazo J .对34个叶绿体基因组的系统发育分析阐明了柑橘属野生种和家养种之间的关系。Mol Biol Evol。2015;32(8):2015–35.<https://doi.org/10.1093/molbev/msv082>。
13. 张x，孙y，Landis JB，Lv Z，Shen J，Zhang H，等.龙胆属(龙胆科)质体系统发育研究:广泛的基因树不一致性及其与质体基因进化速率异质性的关系.BMC Plant Biol。2020;20(1).<https://doi.org/10.1186/s12870-020-02518-w>。
14. 段海，郭杰，宣立，王z，李明，尹勇，等.落羽杉属植物叶绿体基因组的比较研究.BMC基因组公司。2020;21(1).<https://doi.org/10.1186/s12864-020-6532-1>。
15. 粮农组织。[http://www.fao.org/faostat/en/#data](http://www.fao.org/faostat/en/%23data)。2019年访问。
16. 苏伟，王l，雷杰，柴s，刘勇，杨勇，等.基于特定长度扩增片段测序的甘薯群体结构和遗传多样性全基因组评估及核心种质资源的开发(SLAF)。PLoS One。2017;12(2):e172066。
17. 刘d，赵n，翟h，俞x，解q，王l，等.中国主要甘薯栽培种AFLP指纹图谱及遗传多样性研究。J Integr Agric。2012;11(9):1424–33.<https://doi.org/10.1016/S2095-3119(12)60142-7>。
18. 巴西甘薯的遗传多样性。，茄科，旋花科)地方品种。Genet Mol Biol。2008;31(3):725–33.<https://doi.org/10.1590/S1415-47572008000400020>。
19. Wadl PA，Olukolu BA，Branham SE，Jarret RL，Yencho GC，Jackson DM。利用GBSpoly获得的USDA甘薯种质资源的遗传多样性和群体结构。前线工厂科学。2018;9.<https://doi.org/10.3389/fpls.2018.01166>。
20. r包映射工具。<https://cran.r-project.org/web/packages/maptools/index.html>。2020年10月15日访问。
21. r包地图。<https://cran.r-project.org/web/packages/maps/index.html>。2020年10月15日访问。
22. r包ggmap。<https://cran.r-project.org/web/packages/ggmap/index.html>。2020年10月15日访问。
23. r包映射数据。<https://cran.r-project.org/web/packages/mapdata/index.html>。已访问。2020年10月15日。
24. 高等植物的两个最大的叶绿体基因组编码的开放阅读框是必需基因。植物J. 200022(2):97–104.<https://doi.org/10.1046/j.1365-313x.2000.00722.x>。
25. Li M，Yang S，Xu W，Pu Z，Feng J，Wang Z，等.野生甘薯(甘薯近缘野生种Ipomoea trifida)基因组提供了关于贮藏根发育的见解.BMC Plant Biol。2019;19(1).<https://doi.org/10.1186/s12870-019-1708-z>。
26. Isobe S，Shirasawa K，Hirakawa H .甘薯全基因组测序和分析的现状植物细胞代表2019；38(11):1365– 71.<https://doi.org/10.1007/s00299-019-02464-4>。
27. Yoon U，Jeong JC，Kwak S，Yang J，Kim T，Lee H，等.甘薯基因组学研究现状.j植物生物技术醇。2015;42(3):161–7.<https://doi.org/10.5010/JPB.2015.42.3.161>。
28. Amar MH。ycf1-ndhF基因是最有前景的质体基因组条形码，揭示了桃低分类水平的系统发育。J Genet Eng Biotechnol。2020;18(1):42.
29. 董伟，徐c，李c，孙杰，左勇，石s，等.最有前途的陆生植物质体DNA条形码ycf1。Sci Rep. 20155:8348.
30. Nakai M. YCF1:绿色TIC:对de Vries等人评论的回应。植物细胞。2015;27(7):1834–8.<https://doi.org/10.1105/tpc.15.00363>。
31. de Vries J，Sousa FL，Bö lter B，Soll J，Gould SB .YCF1:绿色TIC？植物细胞。2015;27(7):1827–33.<https://doi.org/10.1105/tpc.114.135541>。
32. Bö lter B，Soll J. Ycf1/Tic214对质体蛋白质的积累不是必需的。模塑厂。2017;10(1):219–21.<https://doi.org/10.1016/j.molp.2016.10.012>。
33. 黄杰，孙明。甘薯系列(旋花科)中甘薯及其野生近缘种的遗传多样性和亲缘关系，通过简单序列重复(ISSR)和叶绿体DNA的限制性分析揭示。定理应用遗传算法。2000;100(7):1050–60.<https://doi.org/10.1007/s001220051386>。
34. 杨x，苏w，王l，雷杰，柴s，刘380份甘薯材料的分子多样性和遗传结构的SSR分析。J

Integr Agric。2015;14(4):633–41.[https://doi.org/10.1016/S2095-3119(14](https://doi.org/10.1016/S2095-3119(14)60794-2)

[)60794-2](https://doi.org/10.1016/S2095-3119(14)60794-2)。

1. Tel-Zur N，Abbo S，Myslabodski D，Mizrahi Y .改良的CTAB程序，用于从火龙果属和硒化仙人掌属(仙人掌科)附生仙人掌中分离DNA。Plant Mol Biol报告。1999;17(3):249–54.https://doi.org/10.1023/A:1007656315275.
2. Chen S，Zhou Y，Chen Y，Gu J. Fastp:一种超快速的多功能FASTQ预处理器。生物信息学。2018;34(17):i884–90。<https://doi.org/10.1093/bioinformatics/bty560>。
3. Kim D，Paggi JM，Park C，Bennett C，Salzberg SL。使用HISAT2和HISAT-基因型进行基于图的基因组比对和基因分型。Nat Biotechnol。2019;37(8):907–15.<https://doi.org/10.1038/s41587-019-0201-4>。
4. 甘薯参考叶绿体基因组。[https://www.ncbi.nlm。](http://www.ncbi.nlm/)nih.gov/nu core/NC \_ 026703.1。2020年8月9日访问。
5. Bankevich A，Nurk S，Antipov D，Gurevich AA，Dvorkin M，库利科夫AS，等.一种新的基因组组装算法及其在单细胞测序中的应用.J Comput Biol。2012;19(5):455–77.<https://doi.org/10.1089/cmb.2012.0021>。
6. Marcais G，Delcher AL，Phillippy AM，Coston R，Salzberg SL，Zimin A. MUMmer4:一种快速、通用的基因组比对系统。PLoS Comput Biol。2018;14(1):e1005944。<https://doi.org/10.1371/journal.pcbi.1005944>。
7. MPI-MP氯盒子。<https://chlorobox.mpimp-golm.mpg.de/index.html>。2020年10月10日访问。
8. Tillich M，Lehwark P，佩利泽T，Ulbricht-Jones ES，Fischer A，Bock R，等. GeSeq细胞器基因组的多用途和精确注释。核酸第2017号决议；45(W1):W6–W11。<https://doi.org/10.1093/nar/gkx391>。
9. Greiner S，Lehwark P，Bock r . organilargenomedraw(OGDRAW)第1版。3.1:细胞器基因组图形可视化扩展工具包。核酸第2019号决议；47(W1):W59–64。<https://doi.org/10.1093/nar/gkz238>。
10. 孙杰，董x，曹q，徐t，朱明，孙杰，等.甘薯8个新质体序列的系统比较.2019;7:e6563。<https://doi.org/10.7717/peerj.6563>。
11. 严立，赖x，李x，魏c，谭x，张y .甘薯叶绿体的全基因组及基因表达分析[甘薯]。PLoS One。2015;10:e124083。
12. 野生甘薯叶绿体基因组。http://www . NCBI . nlm.nih.gov/n uccore/？术语=甘薯+叶绿体。2020年10月20日访问。
13. Stothard P，Grant JR，Van DG .使用CGView系列工具可视化和比较环状基因组。简要生物信息。2019;20(4):1576–82.<https://doi.org/10.1093/bib/bbx081>。
14. kea ls M，Moir，Wilson A，Stones-Havas S，Cheung M，Sturrock S，et al . Geneious basic:一个用于序列数据组织和分析的集成和可扩展的桌面软件平台。生物信息学。2012;28(12): 1647–9.<https://doi.org/10.1093/bioinformatics/bts199>。
15. Emms DM，Kelly S. OrthoFinder:比较基因组学的系统发育矫形学推断。Genome Biol。2019;20(1):238.<https://doi.org/10.1186/s13059-019-1832-y>。
16. 价格MN，Dehal PS，Arkin AP。FastTree 2 -用于大型路线的近似最大似然树。PLoS One。2010;5(3):e9490。<https://doi.org/10.1371/journal.pone.0009490>。
17. Huelsenbeck JP，Ronquist F. MRBAYES:系统发育树的贝叶斯推断。生物信息学。2001;17(8):754–5.<https://doi.org/10.1093/bioinformatics/17.8.754>。
18. 无花果树。<http://tree.bio.ed.ac.uk/>。2020年11月10日访问。
19. Langmead B，Salzberg SL。与领结2快速间隙读取对齐。Nat方法。2012;9(4):357–9.<https://doi.org/10.1038/nmeth.1923>。
20. 来自测序数据的SNP调用、突变发现、关联作图和群体遗传参数估计的统计框架。生物信息学。2011;27(21):2987–93.<https://doi.org/10.1093/bioinformatics/btr509>。



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(16): 2078–9.<https://doi.org/10.1093/bioinformatics/btp352>。
2. Danecek P，Auton A，Abecasis G，艾伯斯CA，Banks E，DePristo MA，et al . variant call format and vcf toll .生物信息学。2011;27(15):2156–8.<https://doi.org/10.1093/bioinformatics/btr330>。
3. Wang K，Li M，Hakonarson H. ANNOVAR:来自高通量测序数据的遗传变异函数注释。核酸第2010号决议；38(16):e164。<https://doi.org/10.1093/nar/gkq603>。
4. Alexander DH，Lange K .对用于个体祖先估计的外加剂算法的改进。生物信息学。2011;12(1):246.<https://doi.org/10.1186/1471-2105-12-246>。
5. Purcell S，Neale B，Todd-Brown K，Thomas L，Ferreira MA，Bender D，等.全基因组关联和基于人群的连锁分析的工具集。Am J Hum Genet。2007;81(3):559–75.<https://doi.org/10.1086/519795>。
6. 引物设计程序Primer3的增强和修改。生物信息学。2007;23(10):1289–91.<https://doi.org/10.1093/bioinformatics/btm091>。
7. Untergasser A，cutcutcuta che I，Koressaar T，Ye J，fair Burt BC，Remm M，et al. Primer3 -新功能和接口。核酸第2012号决议；40(15):e115。<https://doi.org/10.1093/nar/gks596>。

# 出版商须知

在已公布的地图和机构附属关系中，Springer Nature对管辖权主张保持中立。