

研究文章开放存取

甘薯品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*）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。结果：对 107 个甘薯品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和结论：对 107 个甘薯品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。关键词：甘薯，甘薯，叶绿体基因组，比较分析，遗传结构

背景

甘薯（*Ipomoea batatas*）是全球重要的粮食作物，被广泛用作工业和生物能源[1]。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, 2]。甘薯属于旋花科甘薯属，是我国唯一的甘薯品种

旋花科六倍体 ($2n = 6x = 90$) 种[3]。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[4–7]。迄今为止，甘薯的起源和进化仍不清楚[8]。现代甘薯被推测为是四倍体祖先和一个

二倍体祖细胞，然后是第二个完整的-

基因组复制[7]。三裂叶甘薯最可能的二倍体祖细胞，这一观点得到了全基因组测序的支持

*通信: cqhe75@yahoo.com

肖士卓、邓毅通对这项工作的贡献是相当的。1 江苏徐州甘薯研究中心/中国农业科学院甘薯研究所，徐州 221131



作者。2021，更正出版物 2021。开放访问本文是根据创作共享属性授权的

4.0 国际许可，允许以任何媒介或形式使用、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，提供到创作共用许可的链接，并表明是否进行了更改。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。

三叶草属植物资料[6]，然而，四倍体祖细胞仍然是未知的。

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, 10]。叶绿体的遗传转化已成为基因工程研究的热点[11]。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, 13]。

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]。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。叶绿体基因组的这些优势更为显著，特别是对于具有复杂核基因组的物种，如甘薯。因此，基于叶绿体基因组的质体组学方法近年来发展迅速[15–18]。

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]。上个世纪，中国广泛种植的甘薯品种主要是日本的‘冲绳100号’和美国的‘南茜·霍尔’及其衍生品种或后代，如徐庶18号，这是中国种植最广泛的甘薯品种，已经成为许多品种的亲本

中国的流行品种[20, 21]。中国甘薯的遗传背景相对狭窄[20]，但随着材料数量的不断增加，有必要对甘薯品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, 22]。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, 23]。然而，由于缺乏高质量的参考基因组，变体调用的可靠性值得商榷。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个甘薯品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个甘薯品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个甘薯品种，其中92个样品来自中国各地(图1桌子, S1)。利用Nova-Seq 6000平台对107个甘薯品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，

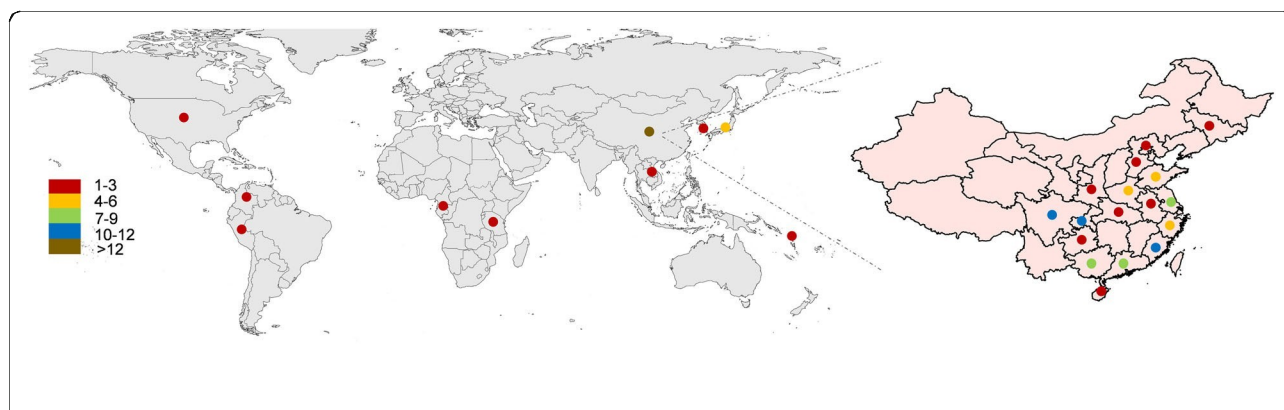


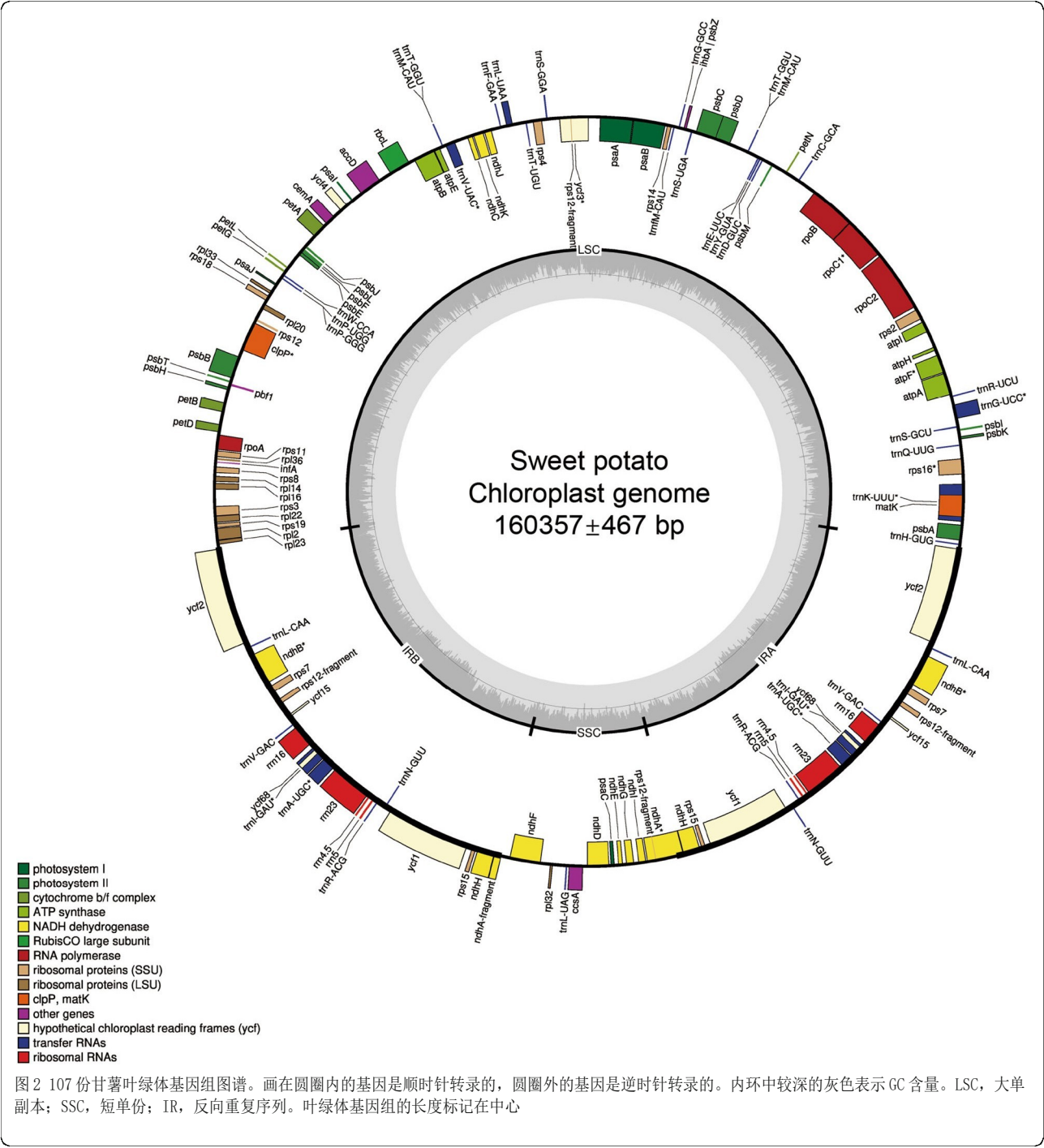
图1 甘薯样品的地理分布和数量。不同的颜色表示从该区域采集的样本数量。左侧是全球样本，右侧是中国样本

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).

选择徐薯 18 的叶绿体基因组作为参考基因组, 筛选比对读片进行组装。重叠群 N50 和 N90 的平均长度分别为 77, 921 bp 和 11, 343 bp, 支架 N50 和 N90 的平均长度

分别达到 85,023 bp 和 16,431 bp, 平均缺口数为 1.88(表 S1). 选择长支架拼接成圆形 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,



888 bp 和 161, 302 bp, 其中 Med = 87, 754 bp, $x = 87, 791, 157$ bp (图 2). LSC (长度 = 87, 589 - 88, 298 bp, Med = 87, 754 bp, $x = 87, 791, 157$ bp) 和 SSC (长度 = 12, 047 - 12, 143 bp, Med = 12, 065 bp, $x = 12, 068, 21$ bp) 被两个 IRs (长度 = 26, 923 - 30, 675 bp, Med = 30, 226 bp, $x = 30, 220, 347$ bp)。甘薯和参考品种徐庶 18 的叶绿体基因组表现出良好的同源性,

这证明了巴塔氏乳杆菌叶绿体基因组的保守性 (图 S1)。

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 个内含子 orf, 由 14 个蛋白编码基因 (ndhA, ndhB,

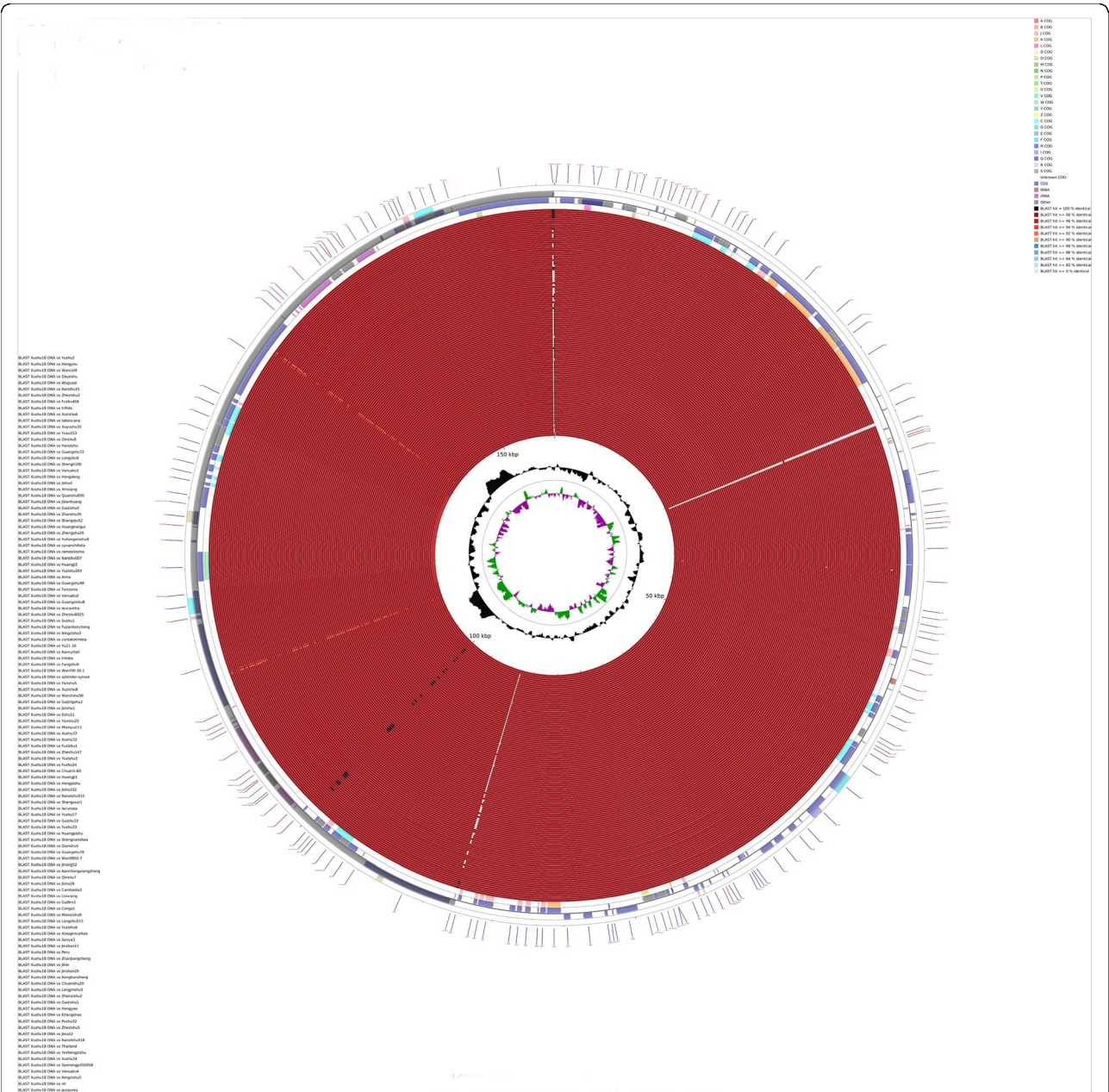


图 3 107 份甘薯与 11 份野生种叶绿体基因组比较分析。特征环 (从最外环开始) 的内容如下: 环 1: 正向链编码序列的 COG 功能类别; 环 2: 正向链序列特征; 环 3: 反向链序列特征; 环 4: 反向链编码序列的 COG 功能类别。接下来的 118 个环显示了通过 BLAST 检测到的序列相似性区域, BLAST 是通过参考基因组和 118 个叶绿体基因组之间的 DNA 序列进行比较的。最里面的两个环分别显示 GC 偏斜和 GC 含量

rps16, *rpoC1*, *ycf3*, *clpP*, *petB*, *petD*, *rpl16*, *atpF* 及两份 *ndhB* 和 *rps*) 和八份 *trnA* (*trnK-UUU*, *trnG-UCC*, *trnL-UAA*, *trnV-UAC* 及两份 *trnI-GAU* 和 *trnA-UGC*) (图. 2).

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 下载并用于与 107 种甘薯进行叶绿体基因组比较分析

栽培品种(图 3). 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, 与其他甘薯品种相比, nil 和紫色甘薯的相似性低于 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

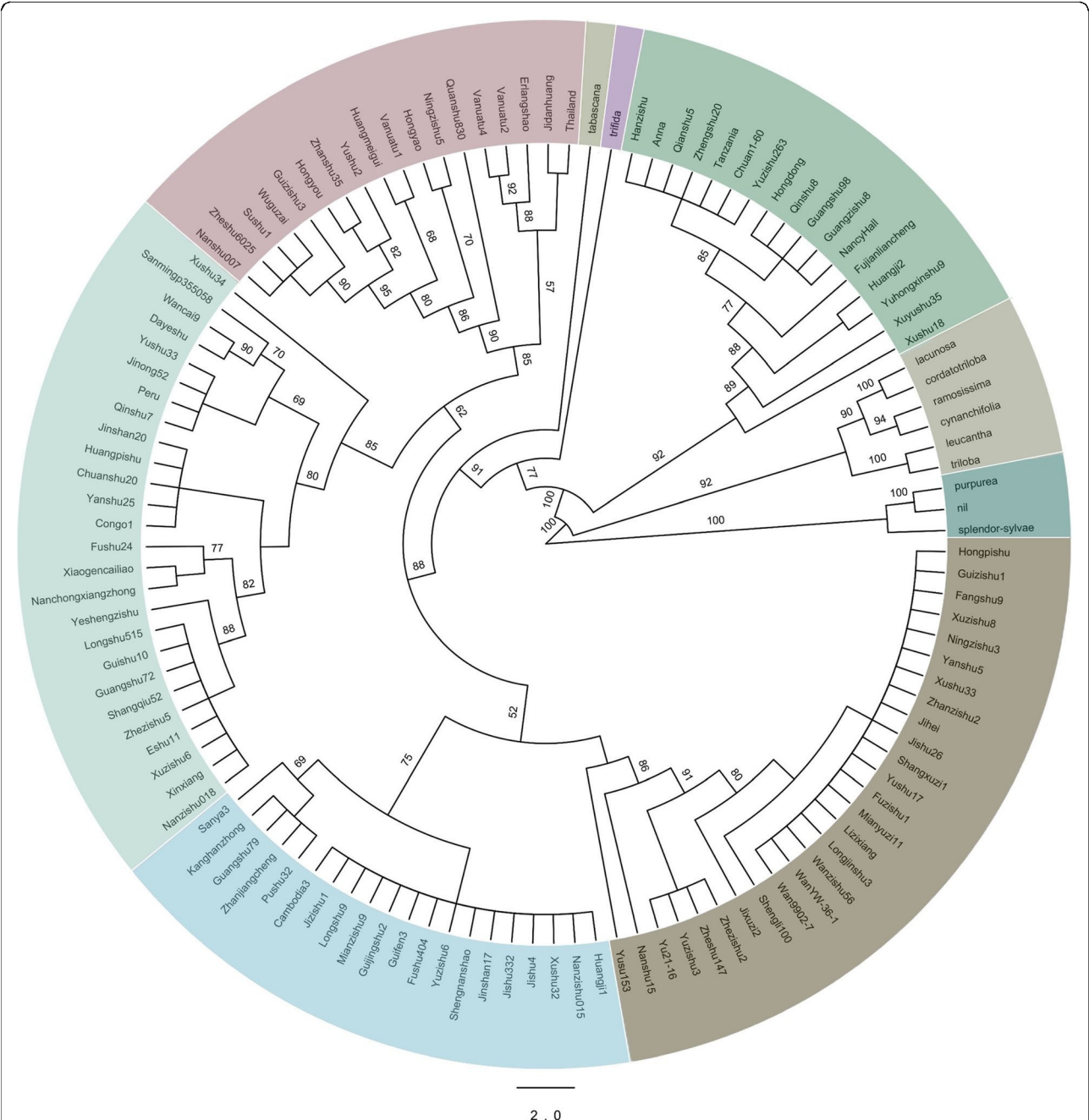


图 4 基于叶绿体单拷贝基因的 107 个甘薯品种和 11 个野生种系统发育树。不同的分支颜色不同, 分支 I、II、IV 和 V 为野生种, 其他为栽培种

Quamoclit 和被认为与甘薯品种有相对远缘关系。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].

考虑到 ycf1 基因的多态性, 从部分品种和 11 个野生种中筛选出 ycf1, 并对氨基酸 (AA) 序列进行比对 (图 S2). 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. 在品种中发现了四个重复序列, 在巴塔塔斯野生种中发现了五个或六个重复序列。此外, 在尼罗罗非鱼中发现了 7 个重复序列, 在紫贻贝中发现了 2 个突变, 因此 ycf1 在品种、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 个甘薯品种进行了系统发育分析 (图. 4). 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 (紫锥菊、紫锥菊和星叶紫锥菊) 离栽培品种最远。第二分支由 6 种野生种 (三叶草属、伦斑蝥属、白薇属、多枝怪柳属、心形怪柳属和腔隙性怪柳属) 组成, 均属于巴达亚组。巴塔塔斯区由甘薯栽培品种的近亲组成。其他两种野生物种 (三叶草和塔巴-卡纳) 被分为两个独立的分支: 分支 IV 和 V, 它们与文化的关系最为密切。二倍体物种三叶草长期以来被认为是甘薯的祖先之一 [6]. 该系统发育分析可能支持这一观点。

徐薯 18 曾是中国种植最广泛的品种, 其亲本之一 ‘南棉 18’ 被归为三号枝。该分枝中的大多数品种与徐薯 18 或其亲本有相对亲缘关系。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 个甘薯品种的两个最大分枝。第七分支主要是橘黄色的

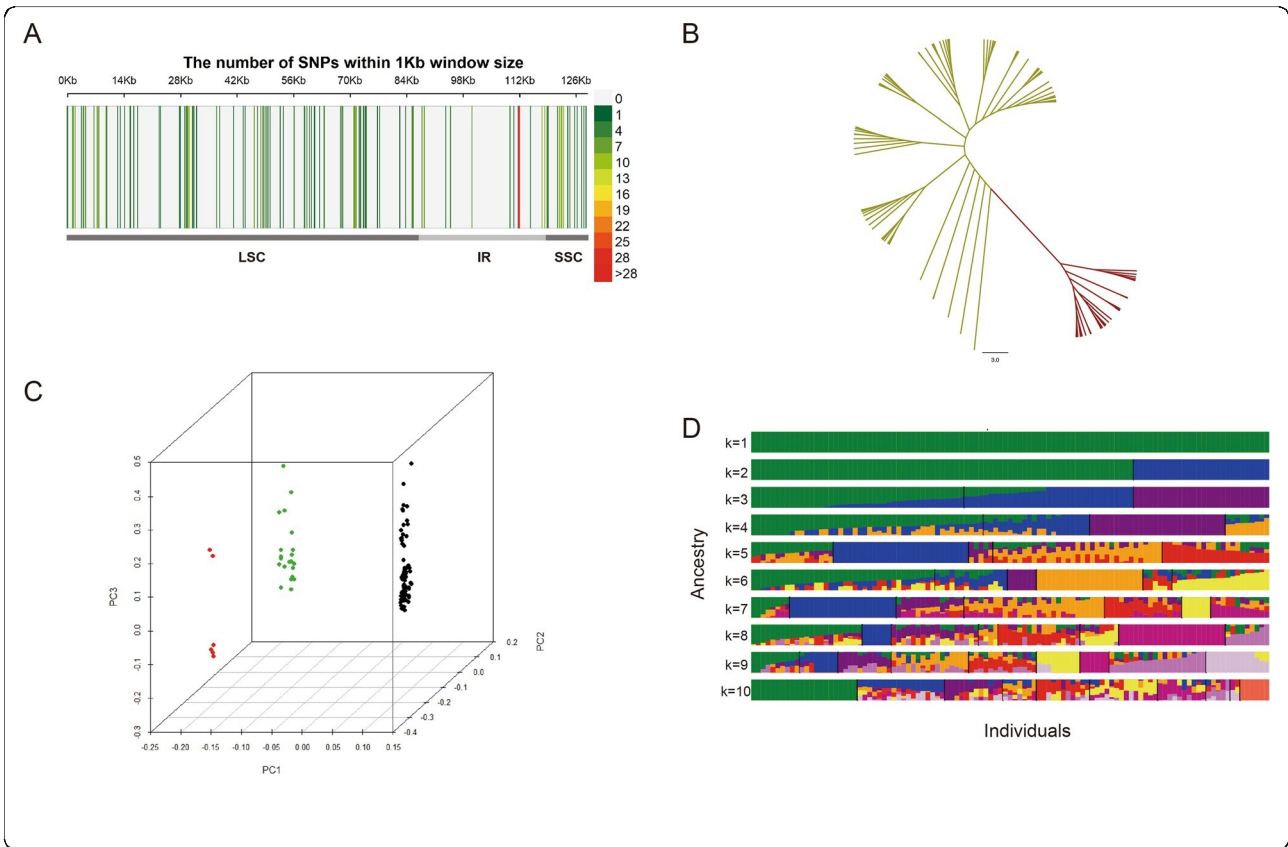


图 5 基于叶绿体基因组变异的 107 个甘薯品种遗传结构分析。叶绿体基因组上 SNP 的密度分布 (一): 系统发育树 (二): 主成分分析 (三): 遗传结构分析 (四): 遗传结构分析

甘薯，占 26 个品种中的 18 个。相比之下，第九分支的大部分品种是紫色甘薯，占 27 个品种中的 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 桌子, S1).

构建了基于 *ycf1* 的系统发育树(图 S3). 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的不同类群。

因此，尽管甘薯品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之间广泛存在的品种间相互引进和杂交所造成的。

变体调用和遗传结构分析

以徐薯 18 叶绿体基因组为对照，检测甘薯 cpDNA 变体。共筛选出 229 个突变位点，其中 118 个 SNPs 和 111 个 InDels(图. 5a)。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 个是非同义变体(表 S2). 该基因 *ycf1* 携带多达 31 个突变位点

对应于比较基因组学分析的结果(图 3).

提取的 SNPs 用于遗传结构分析。进行了系统发育树的构建、主成分分析和种群结构分析(图 5). 系统发生树显示两个主要类群明显聚集(图 5b)。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 5c)。用 1 ~ 10 的 K 值分析种群结构，K = 2 时种群明显分离(图 5d)。K = 2 时，交叉验证(CV)误差也最低(图. S4). 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。总数为设计了 20 对扩增引物。这些引物的长度在 20 bp 至 26 bp 之间，T_m 评分在 57.47° C 至 60.42° C 之间。正向引物和反向引物之间的最大 T_m 差异为 1.95° C。产物的长度大多在 130 bp 至 195 bp 之间，但产物达到 300 bp 的 *Ibcp-15* 除外(表 S3). 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 个甘薯品种，提取其 DNA 作为模板。PCR 结束后进行毛细管电泳。与核 DNA 相比

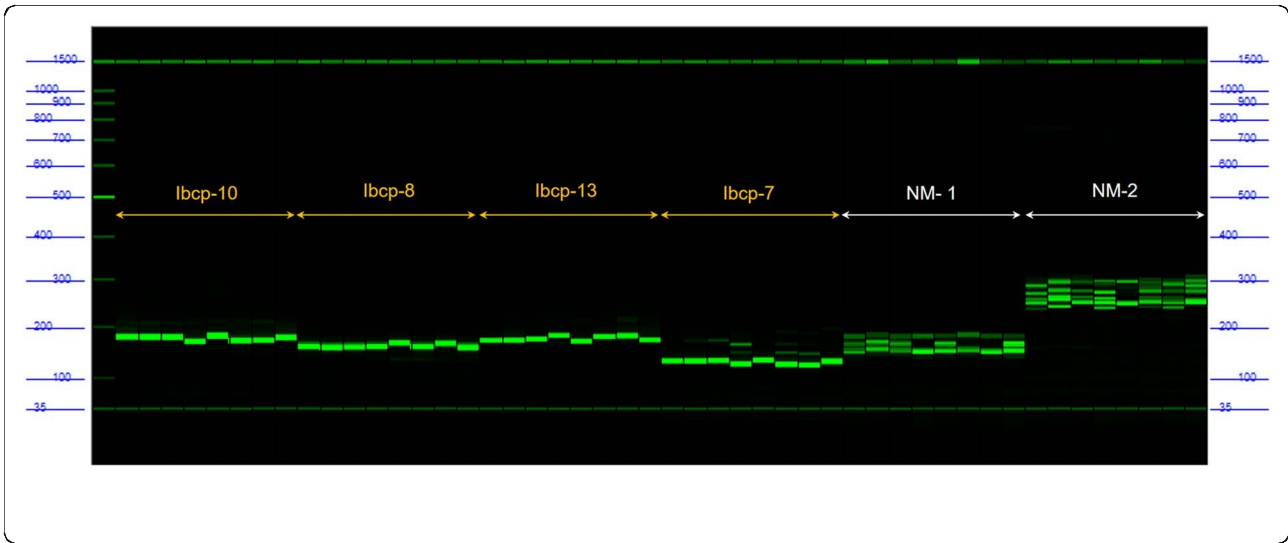


图 6 采用 cpDNA 标记的毛细管电泳。左侧 4 组为 cpDNA 标记扩增产物 (Ibcp-10、Ibcp-8、Ibcp-13、Ibcp-7)，右侧 2 组为

标记时, cpDNA 标记条带单一, 清晰, 可读性更强(图 6). 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品种指纹图谱的构建提供有力的工具。

讨论

本研究对 107 个甘薯品种进行了测序. 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: 三叶虫 [4, 6, 29], *I. 三叶虫* [6], *I. nil* [5], 并且来自栽培品种陈太忠 6 的单倍型基因组 [7]. 由于缺乏高质量的参考基因组, 重测序天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品种徐薯 18 的重新测序将在不久的将来完成 [30, 31]. 使用甘薯的重测序群体进行全基因组关联分析或全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是一个有价值的补充。

石菖蒲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, 尤其是核酮糖二磷酸羧化酶大亚基序列, 相似度大于 98%. 尽管如此, 在 LSC 115 ~ 136 kb 的区域内, 仍可明显区分出 *I. nil* 和 *I. purpureum* (图 3). 这种特征可能是系统发育分析中无刺裸腹蚤和紫色裸腹蚤被划分为单个分支的原因 (图 4).

无论是比较基因组学分析还是变体调用, 都已经证明了 *ycf1* 携带大量的突变 (图 1 和图 2). 3 和 5a, 桌子 S2). 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, 33]. 在本研究中, *ycf1* 不仅在不同品种之间有很大差异, 在不同甘薯品种之间也有很大差异 (图 3). 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 - 36]. 本研究表明, 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 和 S3).

根据叶绿体基因组的单拷贝基因, 将 11 个野生种和 107 个甘薯品种分成 9 个分支 (图 4). 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, 烟粉虱和三叶草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. 甘薯品种的起源长期以来一直不明确, 但普遍认为三叶草是甘薯的祖先之一, 这已被细胞学标记和染色体标记所证实 [8, 37]. 本研究还利用叶绿体比较基因组证实了三叶草是一种与栽培种关系相对密切的野生种 (图 4). 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, 但四倍体烟粉虱在品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品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, 并将这些甘薯品种分为 3 组 [20, 38], 与本研究 PCA 结果一致 (图 5c). 然而, 由叶绿体基因组上的单拷贝基因和变体构建的树之间存在差异 (图 1 和图 2). 4 和 5b) 可能是由叶绿体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). 此外, 部分品种在叶绿体基因组水平上没有得到很好的区分 (图 4), 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. 的放大位置

基因组并不清晰和特异,因此一对引物甚至可以在电泳过程中产生多达20条条带,导致不良的可读电泳结果。本研究根据 InDel calling 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)。cpDNA 标记是分析甘薯遗传多样性或构建甘薯品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),其可被设计为与 sweetpotato 的 cpDNA 标记互补的 dCAP 标记。这些结果将有助于提高我国甘薯品种的同质性。

结论

本研究对 107 个甘薯品种的叶绿体基因组进行了测序、组装和注释。对 107 个甘薯品种和野生种的叶绿体基因组进行了比较分析。甘薯品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。三叶草和烟粉虱与栽培变种关系最为密切,可能参与了甘薯的进化。甘薯品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品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 个甘薯品种,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)。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]。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 attri-布特序列中。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,对发夹环结构的 NEB Next 衔接子进行门控,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。使用 3 μ L USER 酶(NEB, USA),在 PCR 前,在 37° C 下将大小选择的衔接子连接的 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 的配对末端测序。

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]。his T2[41]用于将筛选的读数与参考叶绿体基因组对齐[42]。对齐的读数被用于通过黑桃拼接成支架[43]。支架和参考叶绿体基因组之间的同源性分析是通过 μ m-mer 4.0 进行的[44]。然后选择优质的支架组装成圆形的 DNA 分子。同源性注释在网上进行[45]。GeSeq 模块[46]用于注释圆形 DNA,并手动优化结果。另一个模块 OGDRAW [47]被用来绘制地图。

甘薯比较基因组学分析

11 种野生蕹菜(三裂叶蕹菜、塔巴蕹菜、三裂叶蕹菜、心形蕹菜、白薇蕹菜、星叶蕹菜、多枝蕹菜、白千层蕹菜、腔隙蕹菜、普通蕹菜)叶绿体基因组序列[48]和栽培品种徐薯 18 [49]从 NCBI 下载[50]。将 11 个野生种和 107 个栽培种的基因库文件导入 CGView 软件比较工具[51],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 geneious Basic 4 . 8 . 5 进行比对[52]。

系统发育树的构建

为了识别基因家族, OrthoFinder (v 2.3.14) 管道 [53] 按顺序应用于具有全对全 BLASTP (E 值 $\leq 1e-5$)、互惠最佳命中、通过矫正器和 in-paraolgy 连接的对的十个基因组, 通过矫正器对 E 值和聚类对进行归一化。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)。利用单拷贝直系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 用最大似然和贝叶斯推断 (BI) 构建系统发育树 [54] 和 MrBayes [55]。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 图和 S5)。系统发育树由 Figtree 可视化和修改 [56]。

变体调用

Bowtie2 [57] 用于将干净数据的读数与参考叶绿体基因组对齐。使用 SAMtools 和 BCFtools 执行变量调用 [58, 59]。然后使用 VCFtools 过滤 SNPs 和 InDels [60]。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]。

系统发育树、种群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 误差采用外加剂评估 [62]。默认参数从 $K = 1$ 到 $K = 10$ 。通过 R 包 (条形图) 进行可视化。PCA 由 Plink 进行 [63], 并用 R 软件包 (散点图 3D) 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, 65]。参数设置如下: 产品长度, 小于 200 bp; T_m 评分, $58-64^\circ C$; 正向引物和反向引物之间的 T_m 评分差异, 小于 $2^\circ C$; GC 含量, 35-65%; 引物的长度分布在 20 bp 和 26 bp 之间。由 Sangon Biotech corporation 选择得分最高的引物进行核苷酸序列合成。筛选出 8 个甘薯品种

随机抽取 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 后在美国 AATI 碎片分析仪系统上进行毛细管电泳, 在计算机上读取条带。

缩写

SNP: 单核苷酸多态性; 聚合酶链反应; IR: 反向重复; LSC: 大单份; SSC: 小单份; InDel: 插入并删除; 主成分分析法; cpDNA: 叶绿体 DNACDS: 编码序列; 开放阅读框架

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。品种与徐薯 18 叶绿体基因组的同源性分析
附加文件 2: 图 S2。部分甘薯品种和野生种 $ycf1$ 氨基酸序列的比
附加文件 3: 图 S3。基于 $ycf1$ 的 107 个甘薯和 11 个野生种系统发
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 个甘薯品种的详细资料。
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 促成了变体调用。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 种野生甘薯和徐薯 18 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. tabascan, NC_041207I. 星辉-sylvae, NC_041206I. ramosissima, NC_041205I. cordatotriloba, NC_041204I. 一、白薇, NC_041203I. 腔隙性脑梗死, MH173257I. nil, AP017304I. 紫色, EU118126 徐庶 18, 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
Published online: 13 April 2021

参考

- 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>.
- 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>.
- 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. 2020;7(1). <https://doi.org/10.1038/s41438-020-0314-4>.
- 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>.
- 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>.
- 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>.
- 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>.
- 希里亚克-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>.
- 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>.
- 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>.
- 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>.
- 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](https://doi.org/10.1016/S0076-6879(05)95020-9).
- 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.19.12018.5.001545>.
- 克萊格 MT, 高特 BS, 学习 GJ, 莫頓 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>.
- 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. 2019;6(1). <https://doi.org/10.1038/s41438-019-0191-x>.
- 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>.
- 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>.
- 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 基因组学. 2020;21(1). <https://doi.org/10.1186/s12864-020-6532-1>.
- 粮农组织. <http://www.fao.org/faostat/en/#data>. 2019 年访问。
- 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.
- 刘 d, 赵 n, 翟 h, 俞 x, 解 q, 王 l, 等. 中国主要甘薯品种 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](https://doi.org/10.1016/S2095-3119(12)60142-7).
- 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>.
- 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>.
- r 包映射工具. <https://cran.r-project.org/web/packages/maptools/index.html>. 2020 年 10 月 15 日访问。
- r 包地图. <https://cran.r-project.org/web/packages/maps/index.html>. 2020 年 10 月 15 日访问。
- r 包 ggmap. <https://cran.r-project.org/web/packages/ggmap/index.html>. 2020 年 10 月 15 日访问。
- 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。
- 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. 2000;22(2):97–104. <https://doi.org/10.1046/j.1365-3113.2000.00722.x>.
- Li M, Yang S, Xu W, Pu Z, Feng J, Wang Z, 等. 野生甘薯(三裂叶甘薯)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>.
- 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>.
- 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. 2015;42(3):161–7. <https://doi.org/10.5010/JPB.2015.42.3.161>.
- Amar MH. ycf1-ndhF 基因是最有前景的质体基因条形码, 揭示了桃低分类水平的系统发育. J Genet Eng Biotechnol. 2020;18(1):42.
- 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. 2015;5:8348.
- 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>.
- 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>.
- Bö lter B, Soll J. Ycf1/Tic214 对质体蛋白质的积累不是必需的. 模型工厂. 2017;10(1):219–21. <https://doi.org/10.1016/j.molp.2016.10.012>.
- 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>.
- 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\(14\)60794-2](https://doi.org/10.1016/S2095-3119(14)60794-2).
39. Tel-Zur N, Abbo S, Myslabodski D, Mizrahi Y. 改良的 CTAB 程序, 用于从火龙果属和仙人掌属(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>.
 40. Chen S, Zhou Y, Chen Y, Gu J. Fastp:一种超快速的多功能 FASTQ 预处理程序. 生物信息学. 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
 41. 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>.
 42. 甘薯参考叶绿体基因组. <https://www.ncbi.nlm.nih.gov/nucleotide/106081>. 2020 年 8 月 9 日访问。
 43. 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>.
 44. Marçais 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>.
 45. MPI-MP 氯盒子. <https://chlorobox.mpimp-goim.mpg.de/index.html>. 2020 年 10 月 10 日访问。
 46. 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>.
 47. 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>.
 48. 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>.
 49. 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.
 50. 野生甘薯叶绿体基因组. <http://www.ncbi.nlm.nih.gov/nucleotide/?term=甘薯+叶绿体>. 2020 年 10 月 20 日访问。
 51. Stothard P, Grant JR, Van DG. 使用 CGView 系列工具可视化和比较圆形基因组. 简要生物信息. 2019;20(4):1576–82. <https://doi.org/10.1093/bib/bbx081>.
 52. 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>.
 53. Emms DM, Kelly S. OrthoFinder:比较基因组学的系统发育矫正学推断. Genome Biol. 2019;20(1):238. <https://doi.org/10.1186/s13059-019-1832-y>.
 54. 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>.
 55. 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>.
 56. 无花果树. <http://tree.bio.ed.ac.uk/>. 2020 年 11 月 10 日访问。
 57. Langmead B, Salzberg SL. 与领导 2 快速间隙读取对齐. Nat 方法. 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
 58. 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>.
 59. 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>.
 60. Danecek P, Auton A, Abecasis G, 艾伯斯 CA, Banks E, DePristo MA, et al . variant call format and vcf tool . 生物信息学. 2011;27(15):2156–8. <https://doi.org/10.1093/bioinformatics/btr330>.
 61. 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>.

62. Alexander DH, Lange K. 对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>.
63. 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>.
64. 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>.
65. 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。

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

