

DOI:10.11913/PSJ.2095-0837.2022.10124

郝晨路, 於晓芬, 曲明昊, 赖恩惠, 郭素敏, 高磊. 植物泛基因组研究进展与展望[J]. 植物科学学报, 2022, 40(1): 124-132

Hao CL, Yu XF, Qu MH, Lai EH, Guo SM, Gao L. Current status and prospects of pan-genome studies in plants[J]. *Plant Science Journal*, 2022, 40(1): 124-132

## 植物泛基因组研究进展与展望

郝晨路<sup>1,2</sup>, 於晓芬<sup>1</sup>, 曲明昊<sup>1,2</sup>, 赖恩惠<sup>1,2</sup>, 郭素敏<sup>1</sup>, 高磊<sup>1\*</sup>

(1. 中国科学院武汉植物园, 中国科学院植物物种创新与特色农业重点实验室, 武汉 430074;

2. 中国科学院大学, 北京 100049)

**摘要:** 随着测序技术和生物信息学的快速发展, 已有数百种植物的参考基因组被测序, 极大地促进了植物功能基因组学、进化遗传学和分子育种学等领域的蓬勃发展。然而, 随着研究的深入, 越来越多的证据表明来自单一个体的参考基因组远不能代表整个物种的遗传多样性, 由此催生了泛基因组(Pan-genome)的概念, 并已成功应用于20余种植物的研究, 揭示了丰富的遗传变异, 发掘了大量的新基因, 深化了对相关物种遗传多样性的认识。本文简述了泛基因组的概念、构建方法以及在当前植物研究中的应用现状, 最后对其未来发展进行了展望。

**关键词:** 植物泛基因组; 核心基因; 非必需基因; 结构变异

中图分类号: Q943.2

文献标识码: A

文章编号: 2095-0837(2022)01-0124-09

## Current status and prospects of pan-genome studies in plants

Hao Chen-Lu<sup>1,2</sup>, Yu Xiao-Fen<sup>1</sup>, Qu Ming-Hao<sup>1,2</sup>, Lai En-Hui<sup>1,2</sup>, Guo Su-Min<sup>1</sup>, Gao Lei<sup>1\*</sup>

(1. CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** With the rapid development of sequencing technology and bioinformatics, hundreds of plant reference genomes have been sequenced, which has greatly promoted the development of plant functional genomics, evolutionary genetics, and molecular breeding. However, increasing evidence suggests that one reference from a single individual is insufficient to comprehensively understand the genetic diversity of a given species. The pan-genome concept has been proposed and successfully employed in studies of more than 20 plant species, uncovering abundant novel genes and genetic variations and deepening our understanding of genetic diversity. Here, we briefly introduce the concept and construction of pan-genomes, as well as their applications, development, and prospects in plant research.

**Key words:** Plant pan-genome; Core gene; Dispensable gene; Structural variation

基因组包含了生物体的全套遗传信息, 研究中通常会将一个物种中重要的品系或者最先测出的基因组作为参考基因组, 并以此为基础进行个体或群体水平的遗传变异分析。然而, 由于地域、环境等因素的影响, 同一物种内的不同个体间存在着丰富的遗传变异, 来自单一个体的参考基因组难以覆盖

物种群体的遗传多样性<sup>[1]</sup>。特别是栽培作物与其野生祖先相比, 长期强烈的人工选择导致了严重的遗传瓶颈, 栽培品种的遗传多样性显著降低。而作物的参考基因组常常来自栽培品种, 显然无法代表野生种丰富的遗传变异, 限制了对作物进化和驯化历史的深入研究<sup>[2,3]</sup>。鉴于参考基因组的局限性,

收稿日期: 2021-06-15, 修回日期: 2021-08-27。

基金项目: 国家自然科学基金项目(32170395); 中国科学院植物物种创新与特色农业重点实验室开放课题。

This work was supported by grants from the National Natural Science Foundation of China(32170395) and Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences.

作者简介: 郝晨路(1997-), 女, 硕士研究生, 研究方向为植物基因组学(E-mail: haochenlu19@mails.ucas.ac.cn)。

\* 通讯作者(Author for correspondence. E-mail: leigao@wbcas.cn)。

研究者提出了泛基因组 (Pan-genome) 的概念, 其中前缀“Pan”来自希腊语的“π α ν”, 意思是全部<sup>[4]</sup>。泛基因组自提出以来广泛用于细菌和动植物的研究, 相关的研究报道逐年增长(图 1), 特别是近年来在植物中的应用取得了一系列重要的进展。本文将对泛基因组的概念和构建方法, 及其在植物研究中的现状和发展趋势进行简要概述。

## 1 泛基因组的概念

2005 年, Tettelin 等<sup>[4]</sup>在无乳链球菌(*Streptococcus agalactiae*)的研究中首次提出泛基因组的概念, 包括所有菌株中都存在的核心基因组(Core genome)和部分菌株中特有的非必需基因组(Dispensable genome)(图 2: A)。2007 年, Morgante 等<sup>[5]</sup>首次将这一概念引入植物研究, 但囿于测序技术和成本限制, 并没有很快得到大规模应用。直到 2014 年, Hirsch 等<sup>[6]</sup>构建了玉米(*Zea mays* L.)的泛转录组, Li 等<sup>[7]</sup>和 Schatz

等<sup>[8]</sup>分别报道了野生大豆(*Glycine soja* Sieb. & Zucc.)和水稻(*Oryza sativa* L.)的泛基因组, 才真正开启了植物泛基因组研究的时代。泛基因组中核心基因通常比较保守, 多为管家基因, 能够反映该物种的稳定性; 非必需基因通常具有特定环境适应性, 与个体的特性和防御反应相关, 许多植物的泛基因组研究均发现非必需基因常富集在抗病和防御功能中<sup>[9]</sup>。

## 2 泛基因组构建

### 2.1 样本选择

选择合适的材料构建高质量的基因组, 是开展泛基因组研究的先决条件。泛基因组材料的选择需考虑两个重要因素, 即样本数量和样本特性, 这两个因素直接影响泛基因组的大小、核心和非必需基因的数量与比例, 决定了泛基因组的质量。原则上, 构建泛基因组要尽可能多地增加样本数量, 然而数量越多成本越高, 因此需要选择具有代表性的

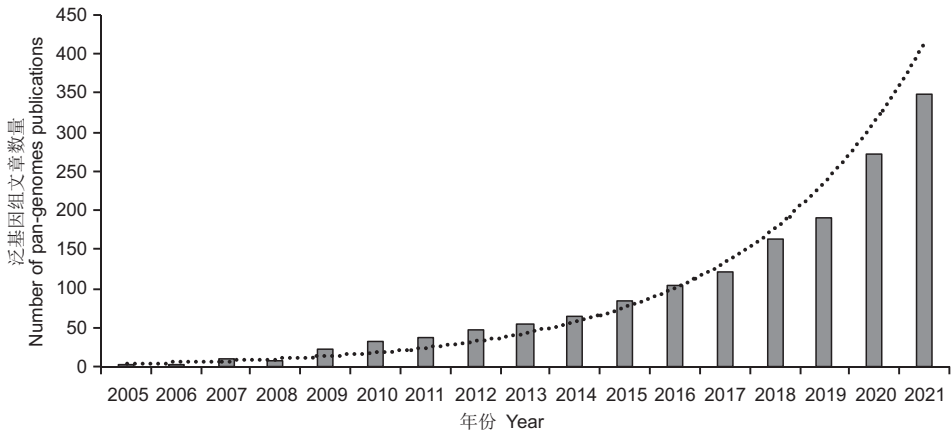


图 1 PubMed 收录的与泛基因组有关的文章  
Fig. 1 Articles related to pan-genomes found in PubMed

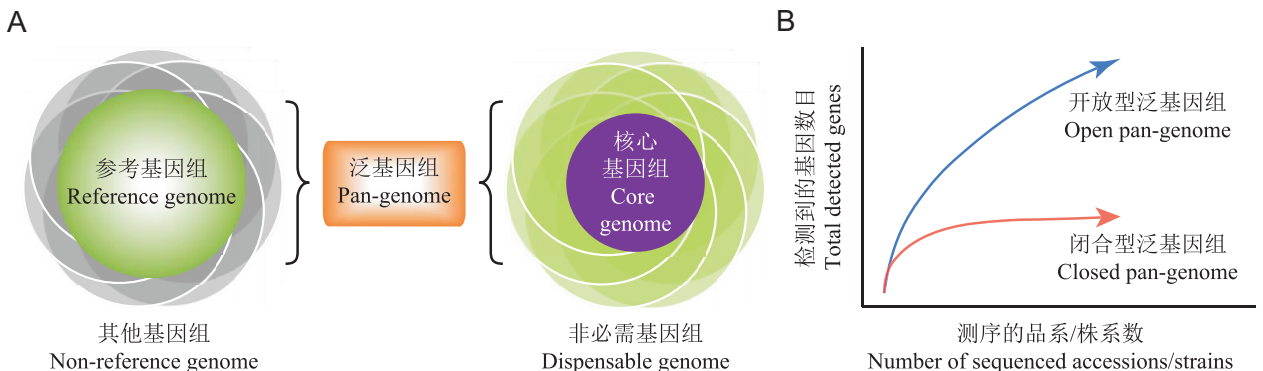


图 2 泛基因组的概念  
Fig. 2 Pan-genome concept

样本。以水稻泛基因组为例, Wang 等<sup>[10]</sup>选取了来源于 89 个国家的 3010 份亚洲稻进行基因组研究, 这些种质代表了全球 780 000 份水稻种质资源中超过 95% 的遗传多样性, 得到了高质量的亚洲稻泛基因组。

栽培作物的驯化改良是一个基因库动态变化的过程。因此, 作物的泛基因组研究加入相关的野生近缘种 (Crop wild relative, CWR), 可以高效地发掘作物驯化过程中丢失的有益基因。如 Gao 等<sup>[11]</sup>对 725 个在地理来源和系统进化方面有代表性的栽培番茄 (*Solanum lycopersicum* L.) 及其野生近缘种进行泛基因组研究, 鉴定到一些野生种质特有的基因, 这些基因与抗病、抗逆、风味、营养等重要性状相关。此外, 向日葵 (*Helianthus annuus* L.)<sup>[12]</sup>、苹果 (*Malus domestica* Borkh.)<sup>[13]</sup>、小麦 (*Triticum aestivum* L.)<sup>[14]</sup> 和大麦 (*Hordeum vulgare* L.)<sup>[15]</sup> 等的泛基因组亦包括了野生近缘种的基因组, 对开发和利用野生资源的基因库具有重要意义。

基于泛基因组大小随测序样本数目的变化趋势, 泛基因组可分为闭合型和开放型两类 (图 2: B)。开放型泛基因组的大小随样本数目的增加而不断增加, 而闭合型泛基因组的大小则随样本数目的增加而增大到一定程度后趋于稳定。泛基因组的开放或闭合特征反映了物种的基因组多样性程度和测序取样的广度, 同时体现了该物种生长环境异质性以及该物种内部遗传物质交流的水平<sup>[16]</sup>。例如大豆<sup>[7, 17]</sup>、玉米<sup>[6]</sup>、番茄<sup>[11]</sup>和甘蓝 (*Brassica oleracea* L.)<sup>[18]</sup> 的泛基因组是闭合型, 可以预测其泛基因组的理论大小。相反, 对于开放型泛基因组, 每个新个体的测序都会持续增加泛基因组的内容, 无法预测最终的泛基因组大小。

## 2.2 构建策略

泛基因组最初提出的时候重点关注基因的存在与缺失变异 (Presence/Absence variations, PAVs)。对于一组注释完善的基因组, 分析每个基因在各个基因组中 PAVs, 可以很容易得到核心基因和非必需基因, 由此得到基于基因的泛基因组 (Gene-based pan-genome) (图 3)。基于基因的泛基因组的明确, 分析方法简单, 在原核生物的泛基因组研究中应用广泛<sup>[19]</sup>。然而, 由于其仅关注基因水平的 PAVs, 而忽略其他序列水平上的

各种变异, 不能全面地反映群体的基因组多样性。因而, 动植物和人类等复杂基因组研究常采用基于序列的泛基因组策略 (Sequence-based pan-genome)<sup>[20]</sup>。

基于序列的泛基因组构建策略, 按提出的时期先后可划分为 4 类: 从头组装 (*De novo assembly*)、迭代组装 (*Iterative assembly*)、Map-to-Pan 和基于变异图组装 (*Graph-based assembly*) (图 3)。从头组装的方法是对每一个体单独测序并从头组装, 再通过多基因组比对分析, 鉴定核心基因、非必需基因和基因组结构变异 (*Structural variations, SVs*)。该策略获得一批高质量的基因组全序列, 开展精细的比较基因组分析, 得到目标物种的泛基因组特征, 但常常并不产生一个非冗余的泛基因组序列, 因而后续的研究难以直接利用。Li 等<sup>[7]</sup>构建的野生大豆、Schatz 等<sup>[8]</sup>和 Zhao 等<sup>[21]</sup>构建的水稻, 以及 Gordon 等<sup>[22]</sup>构建的二穗短柄草 (*Brachypodium distachyon* (L.) P. Beauv.) 等泛基因组均属此类。从头组装构建泛基因组结果相对准确, 但花费的时间较多, 成本也较高。随着测序技术的发展尤其是第三代测序技术的应用, 改善了二代测序读长短的瓶颈问题。其中 PacBio 公司开发的高准确性长读长测序模式显著提高了测序的准确度, 加快了基因组组装的精度和速度。

迭代组装基于物种的参考基因组, 将其他个体的基因组逐个与参考基因组比对, 找出非参考序列, 添加到参考基因组上, 得到一个新的参考基因组, 再进行下一轮比对和添加。通过不断的迭代, 形成该物种的泛基因组。该方法与从头组装相比, 测序成本和计算资源的需求较低, 在早期的植物泛基因组研究中应用颇多, 如甘蓝<sup>[18]</sup>、油菜 (*Brassica napus* L.)<sup>[23]</sup>、木豆 (*Cajanus cajan* (L.) Millsp.)<sup>[24]</sup> 等的泛基因组。

Map-to-pan 同样基于物种的参考基因组进行构建, 与迭代组装的不同之处在于将所有个体的基因组, 分别与参考基因组比对, 鉴定非参考序列; 再把所有的非参考序列合并, 剔除冗余, 得到无冗余的非参考基因组; 把参考基因组和无冗余的非参考基因组合并, 得到泛基因组。该方法流程比较简单, 还可以避免迭代组装因不同迭代顺序可能导致的偏差, 在新近发表的研究中应用较多, 如水

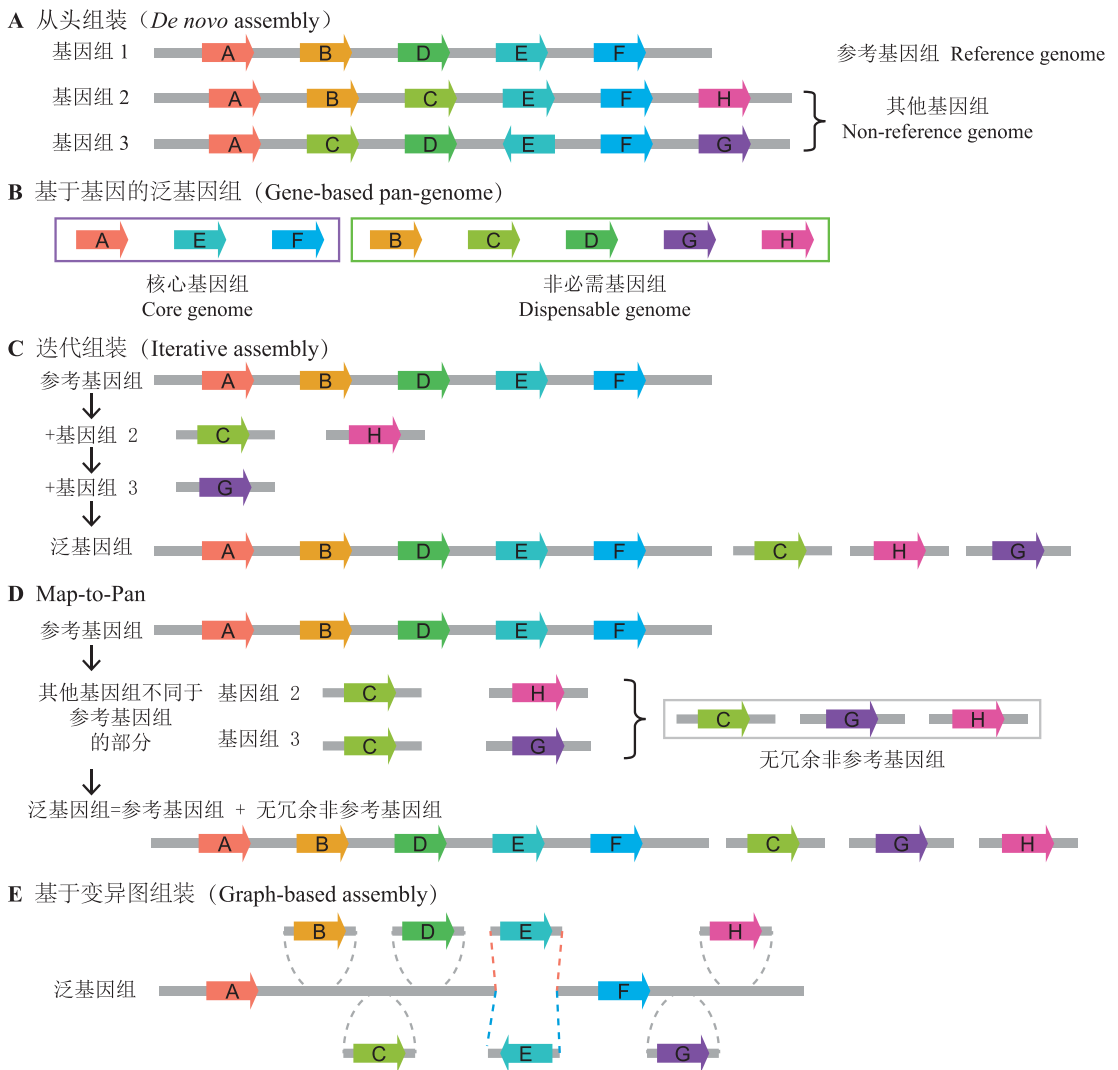


图 3 泛基因组构建策略  
 Fig. 3 Strategies of pan-genome construction

稻<sup>[10]</sup>、番茄<sup>[11]</sup>、苹果<sup>[13]</sup>、大豆 (*Glycine max* (L.) Merr.)<sup>[17]</sup> 和棉花 (*Gossypium hirsutum* L.)<sup>[25]</sup> 等均采用该方法构建泛基因组。

上述 3 种方法得到均是线性泛基因组，无法有效地存储和描述基因组变异。现阶段第三代测序技术迅速发展，可以较低的成本批量产生多个高质量的个体基因组，从而获得大量精确的变异信息。基于变异图组装的方法也采用了从头组装的策略，但是通过多个高质量基因组的精细比较获得变异图，实现基于图形结构的泛基因组构建，能够充分考虑各种形式的遗传变异，完整描述和存储群体的基因组多样性，最近大豆<sup>[26]</sup>、水稻<sup>[27]</sup>、高粱 (*Sorghum bicolor* (L.) Moench)<sup>[28]</sup> 等先后构建了图形化的泛基因组，代表着当前和未来植物泛基因组的主流。

### 3 泛基因组在植物研究中的应用

随着新一代测序技术的飞速发展，迎来了植物泛基因组的大发展。迄今为止，已在 20 余种植物开展了泛基因组研究(表 1)。与基于重测序数据进行变异检测相比，泛基因组研究不仅能够检测出单核苷酸多态性 (Single nucleotide polymorphism, SNP) 和小型插入缺失 (Insertion/deletion, Indel)，还能够大幅提高长片段变异检测的准确性和检出率，从而得到有效的 PAV、拷贝数变异 (Copy number variant, CNV)、大型 Indel、异位 (Translocation) 和倒位 (Inversion) 等各种类型的 SVs，获得物种全面准确的变异信息，有效揭示群体的基因组变异模式。

表 1 已发表的植物泛基因组研究  
Table 1 Summary of published plant pan-genome studies

年份 Year published	物种 Species	样本数量 Number of accessions	构建方法 Approach	核心基因含量/% Percentage of core genes	参考文献 References
2014	玉米 ( <i>Zea mays</i> L.)	503	<i>De novo</i> assembly	39	[6]
	水稻 ( <i>Oryza sativa</i> L.)	3	<i>De novo</i> assembly	92	[8]
	野生大豆 ( <i>Glycine soja</i> Sieb. & Zucc.)	7	<i>De novo</i> assembly	49	[7]
2015	水稻 ( <i>Oryza sativa</i> L.)	1483	<i>De novo</i> assembly	NA	[29]
2016	甘蓝 ( <i>Brassica oleracea</i> L.)	10	Iterative assembly	81	[18]
2017	小麦 ( <i>Triticum aestivum</i> L.)	19	Iterative assembly	64	[30]
	蒺藜苜蓿 ( <i>Medicago truncatula</i> Gaertn.)	15	<i>De novo</i> assembly	33	[31]
	二穗短柄草 ( <i>Brachypodium distachyon</i> (L.) P. Beauv.)	54	<i>De novo</i> assembly	55	[22]
2018	水稻 ( <i>Oryza sativa</i> L.)	3010	Map-to-pan	54~62	[10]
	水稻 ( <i>Oryza sativa</i> L.)	67	<i>De novo</i> assembly	62	[21]
	辣椒 ( <i>Capsicum annuum</i> L.)	383	Map-to-pan	56	[32]
	油菜 ( <i>Brassica napus</i> L.)	53	Iterative assembly	62	[23]
2019	番茄 ( <i>Solanum lycopersicum</i> L.)	725	Map-to-pan	74	[11]
	向日葵 ( <i>Helianthus annuus</i> L.)	493	Map-to-pan	83	[12]
	芝麻 ( <i>Sesamum indicum</i> L.)	5	<i>De novo</i> assembly	58	[33]
	杨属 ( <i>Populus</i> )	10	Iterative assembly	NA	[34]
2020	木豆 ( <i>Cajanus cajan</i> (L.) Millsp.)	89	Iterative assembly	87	[24]
	油菜 ( <i>Brassica napus</i> L.)	8	<i>De novo</i> assembly	56	[35]
	大豆 ( <i>Glycine max</i> (L.) Merr.)	2898	Graph-based assembly	36	[26]
	番茄 ( <i>Solanum lycopersicum</i> L.)	100	PanSV	NA	[36]
	大麦 ( <i>Hordeum vulgare</i> L.)	20	<i>De novo</i> assembly	55	[22]
	小麦 ( <i>Triticum aestivum</i> L.)	15	<i>De novo</i> assembly	74	[14]
	苹果 ( <i>Malus domestica</i> Borkh.)	91	Map-to-pan	81~87	[13]
	拟南芥 ( <i>Arabidopsis thaliana</i> (L.) Heynh.)	7	<i>De novo</i> assembly	80	[37]
2021	狗尾巴草 ( <i>Setaria viridis</i> (L.) P. Beauv.)	598	Map-to-pan	78	[38]
	油菜 ( <i>Brassica napus</i> L.)	1689	Map-to-pan	NA	[39]
	大豆 ( <i>Glycine max</i> (L.) Merr.)	204	Map-to-pan	91	[17]
	棉花 ( <i>Gossypium hirsutum</i> L.)	1913	Map-to-pan	62-86	[25]
	水稻 ( <i>Oryza sativa</i> L.)	33	Graph-based assembly	31	[27]
	高粱 ( <i>Sorghum bicolor</i> (L.) Moench)	16	Graph-based assembly	36	[28]
	芜菁 ( <i>Brassica rapa</i> L.)	18	Graph-based assembly	56	[40]
	茄子 ( <i>Solanum melongena</i> L.)	25	Map-to-pan	88	[41]
	高粱 ( <i>Sorghum bicolor</i> (L.) Moench)	176	Iterative assembly	47	[42]
	玉米 ( <i>Zea mays</i> L.)	26	<i>De novo</i> assembly	27	[43]
	白羽扇豆 ( <i>Lupinus albus</i> L.)	39	Map-to-pan	79	[44]
	萝卜属 ( <i>Raphanus</i> )	11	<i>De novo</i> assembly	65	[45]
	芸薹属 ( <i>Brassica</i> )	246	Iterative assembly	60~79	[46]

### 3.1 泛基因组在植物育种中的应用

基于泛基因组研究可以鉴定大量新基因，为分子设计育种奠定基础(图 4)。长期的驯化选择极大地降低了栽培作物的遗传多样性，因而来自栽培品种的参考基因组可能缺失许多重要基因，如来自水

稻品种‘日本晴’的参考基因组缺失了粒宽和粒重相关基因 *GW5*、耐淹相关基因 *Sub1A* 等，他们均能在泛基因组中找到<sup>[29, 47, 48]</sup>；其他植物的泛基因组亦发现有大量参考基因组缺失的基因，对这些新基因的功能解析将为科学育种提供指导和丰富的基因资源。

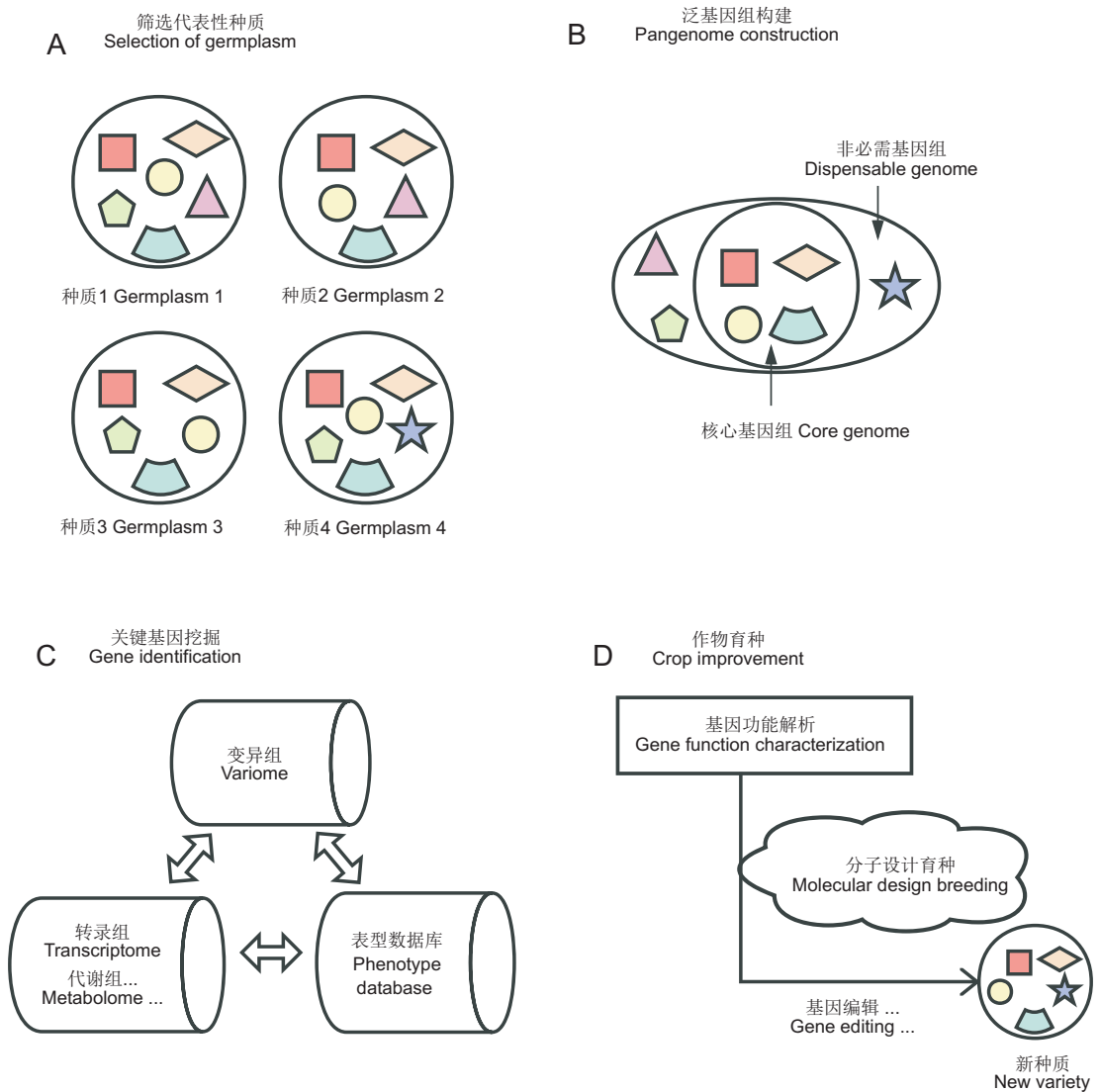


图 4 植物泛基因组及其应用  
Fig. 4 Plant pan-genome and its applications

基于泛基因组进行基因挖掘，可以充分利用多尺度的遗传变异，提高基因定位的准确性。Gao 等<sup>[11]</sup>对番茄的泛基因组研究发现了 *TomLoxC* 基因一个罕见的启动子等位变异，结合对挥发性风味物质开展的、基于 SNPs 的 QTL (Quantitative trait locus) 定位分析，迅速锁定 *TomLoxC* 是影响番茄风味形成的关键基因。

将泛基因组同转录组、代谢组等结合，并利用 QTL 定位和全基因组关联分析 (Genome-wide association study, GWAS) 等策略，可以实现重要性状形成相关基因的快速鉴定 (图 4: C)。Alonge 等<sup>[36]</sup>通过对 100 个番茄品系的泛基因组分析鉴定了 238 490 个 SVs，RNA 测序表明基因表达受到

SVs 的广泛影响，其中一些 SVs 与重要农艺性状相关，包括果实大小、重量和风味品质等。Li 等<sup>[25]</sup>基于棉花泛基因组的分析鉴定到 162 个 QTLs 与纤维品质、产量、开花期等性状相关，其中 84 个是基于泛基因组研究新发现的。

### 3.2 泛基因组在植物进化和驯化方面的应用

进化是指种群里的遗传性状在世代之间的变化，其实质是种群基因频率的改变<sup>[49]</sup>。传统的植物系统发育与分类研究建立在分类群的性状分析基础上，如最初形态性状在进化研究中占重要地位，但由于环境对形态性状的影响较大，使得一些物种的进化地位难以确定<sup>[50]</sup>。而性状的差异实质上是基因型的差异，因此可以通过比较基因组序列的差

异直接研究物种的亲缘和演化关系<sup>[51]</sup>。

构建涵盖不同驯化阶段、不同地域材料的泛基因组，对于探讨作物的进化和驯化具有重要意义。Wang 等<sup>[10]</sup>对 3010 份水稻基因组的分析中，发现粳稻 (*O. sativa* L. ssp. *japonica*) 中缺失很多籼稻 (*O. sativa* L. ssp. *indica*) 特有的等位基因，支持粳稻籼稻独立驯化的假说。Gao 等<sup>[11]</sup>的番茄和 Li 等<sup>[25]</sup>的棉花泛基因组研究发现，驯化和改良中基因均发生了选择性的保留和丢失，而且在驯化过程中丢失的基因显著多于改良过程。Sun 等<sup>[13]</sup>的苹果泛基因组分析结果显示，栽培苹果中约 90% 的基因来自其野生祖先欧洲森林苹果 (*M. sylvestris* (L.) Mill.) 和新疆赛威士苹果 (*M. sieversii* (Ledeb.) M. Roem.)，证实了苹果是沿着丝绸之路从中国传向欧洲，并逐渐驯化为栽培苹果的。综上，泛基因组增加了参考基因组深度和广度，促进了作物进化和驯化的全面解析。

## 4 展望

泛基因组有力地推动了植物学相关研究。基于泛基因组的比较基因组学，能够更好地解析物种和品种形成的分子基础，促进对物种起源和驯化的认识。将泛基因组与 GWAS 和 QTL 分析相结合，可以更高效地鉴别性状形成相关的关键基因。

### 4.1 超级泛基因组

由于作物野生近缘种含有丰富的有益基因，在作物改良中意义重大，Khan 等<sup>[52]</sup>提出了超级泛基因组 (Super-pan-genome) 的概念，即构建一个涵盖了整个属所有物种基因组信息的泛基因组。超级泛基因组超越单一物种的限制，可以揭示多个相关联物种完整的遗传变异，能够促进对 CWR 中特有的抗病抗逆等重要农艺性状关键基因的挖掘，显著扩大栽培作物的可用基因库。此外，超级泛基因组对物种间分化时间和进化事件的解析亦有促进，可为植物进化研究提供新的资源和思路<sup>[52]</sup>。

### 4.2 图形化泛基因组构建及下游分析工具研发

传统的基因组以线性的形式存储于染色体上，无法描述不同个体间的遗传变异。图形化泛基因组将该物种的所有遗传变异信息都整合起来，全面覆盖了多尺度的变异信息，囊括了一个物种全部的遗传信息<sup>[53]</sup>。从线性泛基因组到图形化泛基因组的转变对作物育种尤其重要，作物改良在很大程度上

依赖于遗传变异对农艺性状的影响，图形化泛基因组鉴定到的大量精确的变异信息理论上能够提高 QTL 定位和 GWAS 的准确性<sup>[53, 54]</sup>。

然而，传统的遗传分析工具，包括 GWAS、QTL、转录组分析等，都是基于线性基因组设计的，要利用泛基因组作为参考基因组，原有的软件和算法都需要相应的改进，甚至需要从头开发新的生物信息工具，以适应泛基因组的需求。

### 4.3 局限与挑战

当前植物泛基因组研究取得重大成果的同时也存在一些局限和挑战，首先，泛基因组的组装策略还不够完善，而组装的完整性和准确性是后续分析的前提和保障。其次，基于泛基因组鉴定出了大量遗传变异，如何快速有效地定位性状的关联位点是一个很大的挑战。此外，泛基因组学研究产生了海量的数据，数据的高效存储和可视化也是一个亟待解决的问题。未来测序技术和组装算法的进步将会完善泛基因组的构建，新一代基于泛基因组的 GWAS 和 QTL 等基因定位工具的开发将充分利用泛基因组发掘的多尺度、高密度遗传变异信息，极大地促进植物比较基因组学、功能基因组学和分子育种学的发展。

### 参考文献:

- [1] Yang X, Lee WP, Ye K, Lee C. One reference genome is not enough[J]. *Genome Biol*, 2019, 20(1): 104.
- [2] Olsen KM, Wendel JF. A bountiful harvest: genomic insights into crop domestication phenotypes[J]. *Annu Rev Plant Biol*, 2013, 64(1): 47–70.
- [3] Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification[J]. *Nat Rev Genet*, 2013, 14(12): 840–852.
- [4] Tettelin H, Maignani V, Cieslewicz MJ, Donati C, Medini D, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”[J]. *Proc Natl Acad Sci USA*, 2005, 102(39): 13950–13955.
- [5] Morgante M, De Paoli E, Radovic S. Transposable elements and the plant pan-genomes[J]. *Curr Opin Plant Biol*, 2007, 10(2): 149–155.
- [6] Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, et al. Insights into the maize pan-genome and pan-transcriptome[J]. *Plant Cell*, 2014, 26(1): 121–135.
- [7] Li YH, Zhou G, Ma J, Jiang W, Jin LG, et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits[J]. *Nat Biotechnol*,

- 2014, 32(10): 1045–1052.
- [ 8 ] Schatz MC, Maron LG, Stein JC, Hernandez Wences A, Gurtowski J, *et al.* Whole genome *de novo* assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of *aus* and *indica* [ J ]. *Genome Biol*, 2014, 15(11): 506.
- [ 9 ] Bayer PE, Golicz AA, Scheben A, Batley J, Edwards D. Plant pan-genomes are the new reference [ J ]. *Nat Plants*, 2020, 6(8): 914–920.
- [ 10 ] Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, *et al.* Genomic variation in 3,010 diverse accessions of Asian cultivated rice [ J ]. *Nature*, 2018, 557(7703): 43–49.
- [ 11 ] Gao L, Gonda I, Sun H, Ma Q, Bao K, *et al.* The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor [ J ]. *Nat Genet*, 2019, 51(6): 1044–1051.
- [ 12 ] Hubner S, Bercovich N, Todesco M, Mandel JR, Odenheimer J, *et al.* Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance [ J ]. *Nat Plants*, 2019, 5(1): 54–62.
- [ 13 ] Sun X, Jiao C, Schwaninger H, Chao CT, Ma Y, *et al.* Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication [ J ]. *Nat Genet*, 2020, 52(12): 1423–1432.
- [ 14 ] Walkowiak S, Gao L, Monat C, Haberer G, Kassa MT, *et al.* Multiple wheat genomes reveal global variation in modern breeding [ J ]. *Nature*, 2020, 588(7837): 277–283.
- [ 15 ] Jayakodi M, Padmarasu S, Haberer G, Bonthala VS, Gundlach H, *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding [ J ]. *Nature*, 2020, 588(7837): 284–289.
- [ 16 ] 赵永兵. 泛基因组学分析方法开发及应用 [ D ]. 北京: 中国科学院北京基因组研究所, 2014.
- [ 17 ] Torkamaneh D, Lemay MA, Belzile F. The pan-genome of the cultivated soybean (panSoy) reveals an extraordinarily conserved gene content [ J ]. *Plant Biotechnol J*, 2021, 19(9): 1852–1862.
- [ 18 ] Golicz AA, Bayer PE, Barker GC, Edger PP, Kim H, *et al.* The pangenome of an agronomically important crop plant *Brassica oleracea* [ J ]. *Nat Commun*, 2016, 7(1): 13390.
- [ 19 ] Tettelin H, Medini D. The Pangenome: Diversity, Dynamics and Evolution of Genomes [ M ]. Cham: Springer, 2020: 1–307.
- [ 20 ] Lei L, Goltsman E, Goodstein D, Wu GA, Rokhsar DS, *et al.* Plant pan-genomics comes of age [ J ]. *Annu Rev Plant Biol*, 2021, 72(1): 411–435.
- [ 21 ] Zhao Q, Feng Q, Lu H, Li Y, Wang A, *et al.* Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice [ J ]. *Nat Genet*, 2018, 50(2): 278–284.
- [ 22 ] Gordon SP, Contreras-Moreira B, Woods DP, Des Marais DL, Burgess D, *et al.* Extensive gene content variation in the *Brachypodium distachyon* pan-genome correlates with population structure [ J ]. *Nat Commun*, 2017, 8(1): 2184.
- [ 23 ] Hurgobin B, Golicz AA, Bayer PE, Chan CK, Tirnaz S, *et al.* Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus* [ J ]. *Plant Biotechnol J*, 2018, 16(7): 1265–1274.
- [ 24 ] Zhao J, Bayer PE, Ruperao P, Saxena RK, Khan AW, *et al.* Trait associations in the pangenome of pigeon pea (*Cajanus cajan*) [ J ]. *Plant Biotechnol J*, 2020, 18(9): 1946–1954.
- [ 25 ] Li J, Yuan D, Wang P, Wang Q, Sun M, *et al.* Cotton pan-genome retrieves the lost sequences and genes during domestication and selection [ J ]. *Genome Biol*, 2021, 22(1): 119.
- [ 26 ] Liu Y, Du H, Li P, Shen Y, Peng H, *et al.* Pan-genome of wild and cultivated soybeans [ J ]. *Cell*, 2020, 182(1): 162–176.
- [ 27 ] Qin P, Lu H, Du H, Wang H, Chen W, *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations [ J ]. *Cell*, 2021, 184(1): 1–17.
- [ 28 ] Tao Y, Luo H, Xu J, Cruickshank A, Zhao X, *et al.* Extensive variation within the pan-genome of cultivated and wild sorghum [ J ]. *Nat Plants*, 2021, 7(1): 766–773.
- [ 29 ] Yao W, Li G, Zhao H, Wang G, Lian X, *et al.* Exploring the rice dispensable genome using a metagenome-like assembly strategy [ J ]. *Genome Biol*, 2015, 16(1): 187.
- [ 30 ] Montenegro JD, Golicz AA, Bayer PE, Hurgobin B, Lee H, *et al.* The pangenome of hexaploid bread wheat [ J ]. *Plant J*, 2017, 90(5): 1007–1013.
- [ 31 ] Zhou P, Silverstein KA, Ramaraj T, Guhlin J, Denny R, *et al.* Exploring structural variation and gene family architecture with *de novo* assemblies of 15 *Medicago* genomes [ J ]. *BMC Genomics*, 2017, 18(1): 261.
- [ 32 ] Ou L, Li D, Lv J, Chen W, Zhang Z, *et al.* Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses [ J ]. *New Phytologist*, 2018, 220(2): 360–363.
- [ 33 ] Yu J, Golicz AA, Lu K, Dossa K, Zhang Y, *et al.* Insight into the evolution and functional characteristics of the pangenome assembly from sesame landraces and modern cultivars [ J ]. *Plant Biotechnol J*, 2019, 17(5): 881–892.
- [ 34 ] Zhang B, Zhu W, Diao S, Wu X, Lu J, *et al.* The poplar pangenome provides insights into the evolutionary history of the genus [ J ]. *Commun Biol*, 2019, 2(1): 215.
- [ 35 ] Song JM, Guan Z, Hu J, Guo C, Yang Z, *et al.* Eight



- high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus* [J]. *Nat Plants*, 2020, 6(1): 34–45.
- [36] Alonge M, Wang X, Benoit M, Soyk S, Pereira L, *et al.* Major impacts of widespread structural variation on gene expression and crop improvement in tomato [J]. *Cell*, 2020, 182(1): 145–161.
- [37] Jiao WB, Schneeberger K. Chromosome-level assemblies of multiple *Arabidopsis* genomes reveal hotspots of rearrangements with altered evolutionary dynamics [J]. *Nat Commun*, 2020, 11(1): 989.
- [38] Mamidi S, Healey A, Huang P, Grimwood J, Jenkins J, *et al.* A genome resource for green millet *Setaria viridis* enables discovery of agronomically valuable loci [J]. *Nat Biotechnol*, 2020, 38(10): 1203–1210.
- [39] Song JM, Liu DX, Xie WZ, Yang Z, Guo L, *et al.* BnPIR; *Brassica napus* pan-genome information resource for 1689 accessions [J]. *Plant Biotechnol J*, 2021, 19(3): 412–414.
- [40] Cai X, Chang L, Zhang T, Chen H, Zhang L, *et al.* Impacts of allopolyploidization and structural variation on intraspecific diversification in *Brassica rapa* [J]. *Genome Biol*, 2021, 22(1): 166.
- [41] Barchi L, Rabanus-Wallace MT, Prohens J, Toppino L, Padmarasu S, *et al.* Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding [J]. *Plant J*, 2021, 107(2): 579–596.
- [42] Ruperao P, Thirunavukkarasu N, Gandham P, Selvanayagam S, Govindaraj M, *et al.* Sorghum pan-genome explores the functional utility for genomic-assisted breeding to accelerate the genetic gain [J]. *Front Plant Sci*, 2021, 12(1): 666342.
- [43] Hufford MB, Seetharam AS, Woodhouse MR, Chougule KM, Ou S, *et al.* *De novo* assembly, annotation, and comparative analysis of 26 diverse maize genomes [J]. *Science*, 2021, 373(6555): 655–662.
- [44] Hufnagel B, Soriano A, Taylor J, Divol F, Kroc M, *et al.* Pangenome of white lupin provides insights into the diversity of the species [J]. *Plant Biotechnol J*, 2021, 19(12): 2532–2543.
- [45] Zhang X, Liu T, Wang J, Wang P, Qiu Y, *et al.* Pan-genome of *Raphanus* highlights genetic variation and introgression among domesticated, wild, and weedy radishes [J]. *Mol Plant*, 2021, 14(12): 2032–2055.
- [46] Bayer PE, Scheben A, Golicz AA, Yuan Y, Faure S, *et al.* Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids [J]. *Plant Biotechnol J*, 2021, 19(12): 2488–2500.
- [47] Weng J, Gu S, Wan X, Gao H, Guo T, *et al.* Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight [J]. *Cell Res*, 2008, 18(12): 1199–1209.
- [48] Xu K, Xu X, Fukao T, Canlas P, Maghirang-Rodriguez R, *et al.* *Sub1A* is an ethylene-response-factor-like gene that confers submergence tolerance to rice [J]. *Nature*, 2006, 442(7103): 705–708.
- [49] Laland K, Uller T, Feldman M, Sterelny K, Müller GB, *et al.* Does evolutionary theory need a rethink? [J]. *Nature*, 2014, 514(7521): 161–164.
- [50] 石开明, 彭昌操, 彭振坤, 罗正荣. DNA 序列在植物系统进化研究中的应用 [J]. 湖北民族学院学报 (自然科学版), 2002, 20(4): 5–10.
- [51] 丁士友, 张春林, 顾红雅, 陈章良. DNA 水平上的植物系统学研究进展 [J]. 西北植物学报, 1996, 4: 446–456.
- Ding SY, Zhang CL, Gu HY, Chen ZL. Progresses of the studies on plant systematics at DNA level [J]. *Acta Botanica Boreali-Occidentalia Sinica*, 1996, 4: 446–456.
- [52] Khan AW, Garg V, Roorkiwal M, Golicz AA, Edwards D, *et al.* Super-pangenome by integrating the wild side of a species for accelerated crop improvement [J]. *Trends Plant Sci*, 2020, 25(2): 148–158.
- [53] Tao Y, Jordan DR, Mace ES. A graph-based pangenome guides biological discovery [J]. *Mol Plant*, 2020, 13(9): 1247–1249.
- [54] Tao Y, Jordan DR, Mace ES. Crop genomics goes beyond a single reference genome [J]. *Trends Plant Sci*, 2019, 24(12): 1072–1074.

(责任编辑: 周 媛)