

泛基因组学在植物中的应用研究进展

王灿^{1,2}, 王艳芳¹, 张应华¹, 许俊强^{1*}

(1.云南农业大学 云南省滇台特色农业产业化工程研究中心,云南 昆明 650201;

2.云南省文山州农业科学院,云南 文山 663000)

摘要:随着DNA测序技术的发展,越来越多的物种基因组被测序,即使是在同一物种内,基因组也存在着较大差异.为了更全面的在基因组水平上探究个体间的遗传多样性,提出泛基因组的概念,并广泛应用于细菌、真菌、动植物等.本文对泛基因组在植物中的应用研究进行系统的综述,以期对泛基因组在植物中的深入研究提供参考.

关键词:泛基因组;植物;核心基因组

中图分类号:Q943

文献标志码:A

文章编号:2095-7300(2020)02-0051-08

Research Progress on the Applications of Pan-genomics in Plants

Wang Can^{1,2}, Wang Yanfang¹, Zhang Yinghua¹, Xu Junqiang^{1*}

(1.Dian-Tai Engineering Research Center for Characteristic Agriculture Industrialization of Yunnan Province, Yunnan Agricultural University, Kunming 650201, China;

2.Wenshan Academy of Agricultural Sciences of Yunnan Province, Wenshan 663000, China)

Abstract: With the development of DNA sequencing technology, more and more species' genomes have been sequenced. Even within the same species, there are great differences in the genomes. In order to explore the genetic diversity among individuals at the level of genome more comprehensively, the concept of pan-genome was proposed, which was widely used in bacteria, fungi, plants and animals, etc. In this paper, the application of pan-genome was reviewed in order to provide reference for the further studies of pan-genome in plants.

Keywords: pan-genome; plants; core genome

1995年,第一株流感嗜血杆菌(*Haemophilus influenzae* Rd KW20)全完整基因组基因测序图谱并组

装^[1],由此,细菌全基因组测序开始兴起.随后,詹氏甲烧球菌(*Methanococcus jannaschii*)^[2]、大肠杆菌K-12

收稿日期:2019-04-09

基金项目:国家自然科学基金(31560560);云南省应用基础研究计划项目(2015FD019)

作者简介:*通信作者,许俊强,博士,讲师,研究方向:蔬菜育种与分子生物学,E-mail:xujunqiang101@163.com.王灿,E-mail:1561848631@qq.com

引文格式:王灿,王艳芳,张应华,等.泛基因组学在植物中的应用研究进展[J].湖南生态科学学报,2020,7(2):51-58.Wang C, Wang Y F, Zhang Y H, et al. Research progress on the applications of pan-genomics in plants[J]. Journal of Hunan Ecological Science, 2020, 7(2): 51-58.

(*Escherichia coli* K-12)^[3]等细菌的全基因组相继被测序. 2005年, 微生物泛基因组概念^[4] (Pan-genome) 由 Tettelin 提出. 泛基因组研究不仅可以全面对某一物种内遗传多样性进行研究分析, 而且可以探究不同个体间的进化关系^[5-6]. 在微生物方面还可以对菌种进化、适应性及群体结构进行研究分析, 同时, 在重要毒力因子发现、新疫苗设计等方面也有巨大的挖掘潜力和应用价值^[7-9]. 2009年, Li 等^[10]初次采用全新基因组组装方法对多个人类个体基因组进行拼接, 发现了个体中独有的 DNA 序列和功能基因, 并首次提出了“人类泛基因组”的概念; 2013年泛基因组测序开始应用于动植物研究领域, 目前泛基因组研究中已涉及细菌、真菌、病毒、动物、植物、人类等. 因此, 泛基因组学越来越成为研究的热点.

1 泛基因组学基本概念

泛基因组是对某一物种全部基因的总称, 由核心基因组、非必须基因组及特异性基因组组成^[11-12]. 核心基因组指在所有株系中都存在的基因^[13]; 非必须基因组是指仅在部分株系中存在的基因^[5, 14]; 特异性基因组指仅在某一个株系存在的基因^[15-16], 如图 1 所示.

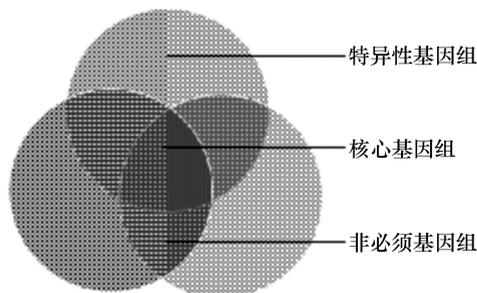


图 1 泛基因组组成

Fig. 1 Pan-genome composition

核心基因组中大多数是必须基因和信息基因, 由所有样本中共同拥有的序列组成, 一般与物种生物学功能和主要表型特征相关, 负责该物种基本生物学功能和主要的表型特征, 如核孔的调节转运、基因的复制、转录和翻译, 反映了物种的稳定性^[13, 17-18]; 非必须基因不是所有株系中共有的基因, 它由仅在单个样本或部分样本中存在的序列组成, 它体现了株系间的差异性和多样性, 反映了物种的特性, 具有选择优势, 如对特定环境的适应性、抗生

素抗药性和对新寄主的寄生性等^[4, 13, 19]; 特异性基因是仅在某一个株系中所特有的基因^[4].

随着测序的基因组数目增加, 物种的泛基因组大小增加到一定程度后收敛于某一值, 并且完全足够描述这一物种, 称为闭合型泛基因组 (Close Pan-genome). 例如, 对 8 株独立的炭疽芽孢杆菌 (*Bacillus anthracis*) 研究发现, 在测序了 4 个基因组之后, 泛基因组大小收敛于某一值^[4]. 反之, 某些物种随着测序基因组数目的增加, 物种的泛基因组大小也不断增加, 这些物种的泛基因组则称为开放型泛基因组 (Open Pan-genome). 测序了 8 个链球菌 (Group B *Streptococcus*) 菌株, 其泛基因组包含 2 713 个基因, 其中有 1 086 个核心基因和 907 个非必须基因, 平均每测序一个基因组, 其泛基因组集合增加 33 个新基因^[15]. 分析 5 个酿脓链球菌 (*Streptococcus pyogenes*) 菌株中也发现类似的情况, 每测序一个基因组, 泛基因组集合增加 27 个新基因^[15].

2 泛基因组在植物研究领域中的应用

2.1 大豆

邱丽娟等^[12]对 7 份代表性野生大豆 (中国北方、黄淮、南方和东北地区、日本、韩国和俄罗斯) 进行测序与独立 *De Novo* 组装, 构建了大豆泛基因组. 结果发现野生大豆中约 48.6% 为核心基因, 51.4% 为非核心基因, 主要与野生大豆抗耐性和非生物逆境相关, 决定着野生大豆对环境的适应性, 这为作物改良, 导入栽培大豆提供了候选基因.

2.2 花生

花生是我国重要的经济作物, 是生产食用植物油的原料. 为破译异源四倍体野生花生基因组, 殷冬梅等^[20]对 1 个异源四倍体野生花生 (*Arachis monticola*) 进行基因组组装, 采用单分子实时测序 (Single Molecule Real Time, SMRT) 得到高质量野生花生参考基因组, 大小为 2.62 Gb. 研究员对其基因组进行区分得到两个亚基因组, 并与祖先 *A. duranensis* 和 *A. ipaensis* 基因组比对, 发现野生花生 (*A. monticola*) 的两个亚基因组与其分别对应的祖先基因组大小相近, 这为后续的研究打下了坚实的基础. 栽培花生 (*A. hypogaea* Linn), 异源四倍体植物 (AABB) 具有两个亚基因组, 是两个野生花生物种 (*A. duranensis*

和 *A. ipaensis*) 的自然杂种. Bertoli 等^[21] 对其二倍体祖先花生 *A. duranensis* 和 *A. ipaensis* 基因组测序, 分别得到 1 211 Mb 和 1 512 Mb 基因组, 并发现它们两个的基因组与栽培花生 (*A. hypogaea* Linn) 的 A 和 B 两个亚基因组序列极其相近, 覆盖了其基因组的 96%. 因此可以对鉴定抗病基因、栽培花生转录组组装以及研究栽培花生 A 和 B 两个亚基因组的遗传重组等重要问题提供帮助, 为选育高产、适应性强、抗病虫害、抗逆境的花生品种奠定基础. 同时其基因组的破译也了解花生属及豆科作物进化提供了重要的科学参考价值, 为促进花生以及其他油料作物的功能基因组学发展奠定基础.

2.3 水稻

张建伟等^[22] 对代表籼稻亚种的两个主要品种群, 珍汕 97 和明恢 63 进行了基因组测序. 找到了两个籼稻亚种间各自特异的基因及相同的基因, 为研究水稻的遗传多样性, 及一系列农业性状解读提供分子基础. 黄学辉等^[23] 和 Stein 等^[24] 分别对 66 个水稻材料的泛基因组数据和 13 个水稻品种的基因组数据进行分析发现: 在水稻各材料基因组中存在广泛的插入与缺失, 转座子 (Transposable Element, TE) 的选择性扩增和缺失在其中起着关键的作用. 该泛基因组数据将会进一步为水稻的进化功能研究和复杂性状潜在的目标位点精确定位提供参考.

2.4 辣椒

Kim 等^[25] 报道了 2 个驯化品种辣椒 (*Capsicum baccatum* 和 *C. chinense*) 的基因组图谱, 并对已测序的辣椒 (*C. annuum*) 基因组进行了优化. 对比分析 3 种辣椒基因组发现, 在染色体 3、5、9 中有涉及易位的动态基因组重排. 通过不同辣椒基因组的长末端重复序列反转录转座子 (Long Terminal Repeat-Retrotransposons, LTR-Rs) 插入模式的研究, 发现在基因组中有特定 LTR-Rs 的积累, 以及 LTR-Rs 对新候选基因产生和辣椒基因组的多样化起到促进作用. 通过对基因和重复序列进行比对后发现, 加倍的核苷酸结合和富含亮氨酸的重复序列 (Nucleotide-binding and Leucine-rich-Repeat, NLR) 抗病基因家族基因在被子植物中广泛存在. 指出辣椒基因组中具有大量 NLRs (包括功能性抗病基因), LTR-R 驱动的逆转录复制在其中发挥了关键作用, 同时也表明了至少有 5%~18% 的植物 NLRs 是由 LTR-R 驱动的逆转录复制产生的. 此外, 欧立军等^[26] 为研究辣

椒种内的基因存在 - 缺失变异 (Presence-Absence Variation, PAV) 对不同种之间的性状差异影响. 通过对 383 份辣椒材料进行 *De novo* 测序和与 *C. annuum* L. 栽培品种 Zunla-1 基因组的比对建立了辣椒的泛基因组; 同时利用 PAV 结果对全基因组关联分析 (Genome-wide association study, GWAS) 和基因组进化分析, 结果发现在果实呈现黄色或橙色的辣椒中辣椒红素合成酶 *Ces* (*Capsanthin/capsorubin synthase*) 基因附近区域有明显缺失; 另外在 (*Pun1*) region 内及其附近有基因缺失的辣椒, 其果实中辣椒素含量明显下降. 该研究成果不仅揭示了辣椒种间基因组遗传变异的多样性, 同时也为基因 PAV 在对基因组进化分析和基因组关联研究中奠定了基础和提供了见解.

2.5 南瓜

Sun 等^[27] 通过对 2 种南瓜 (*Cucurbita maxima* 和 *C. moschata*) 基因组进行 *De Novo* 组装, 得到高质量基因组序列 271.4 Mb 和 269.9 Mb. 测序结果表明在基因组序列中 40% 以上为重复序列, 在这些重复序列中, *C. maxima* 中的 69.9% 和 *C. moschata* 中的 62.9% 被注释为长末端重复序列 (Long Terminal Repeat, LTR). 根据贝宁部落 (*Benincaseae* tribe) 中甜瓜、黄瓜和西瓜中不同遗传距离 (Genetic distance) 将基因组分成两个同源亚基因组, 结果发现这两个亚基因组保留了相似数量的基因, 这为南瓜中异源四倍体化事件提供了证据, 同时也提供了多倍体基因组进化的见解和葫芦科作物遗传改良的宝贵资源.

2.6 烟草

Xu 等^[28] 对 2 种野生烟草 *Nicotiana attenuata* (2.5 Gb) 和 *N. obtusifolia* (1.5 Gb) 的基因组进行了测序和组装, 并将基因组进化与烟草中防御代谢物 - 尼古丁的两种合成途 (多胺和烟酰胺腺嘌呤二核苷酸) 联系起来. 揭示了在茄科全基因组三倍化事件 (Whole Genome Triplication, WGT) 后, 一系列快速扩增的转座因子 (TE) 导致其基因组膨胀, 促进了重复基因之间的表达差异, 同时也揭示了烟草诱导的信号传导和防御的进化, 为 TE 和基因重复促进与植物健康相关的关键代谢创新的出现提供了依据, 同时进一步表明基因重复和转座因子的插入在某些特定的代谢生物合成途径进化中相互作用的重要性.

2.7 向日葵

向日葵 (*Helianthus annuus* L.) 是菊科向日葵属,原产于南美洲,中国、欧洲、北美洲均有分布,是一种能在干旱条件下保持稳产的油料作物。Badouin 等^[29]采用 PacBio RS II 测序平台 SMRT 测序得到大小为 3 Gb 的基因组,发现超过四分之三的向日葵基因组由长末端重复序列反转录转座子 LTR-RTs 组成,同时研究者将向日葵、生菜、朝鲜蓟、咖啡和外群物种葡萄进行比对分析,表明向日葵、朝鲜蓟、生菜都经历了一次全基因组三倍化事件(WGT)另外向日葵又经历特异性的全基因组复制,最终形成 17 条染色体组。Hübner 等^[30]对栽培种品系 287 个、美国原地方品种 17 个、野生近缘种 189 个共 493 份向日葵种质资源进行测序。结果得到 61 205 个基因,发现其中大约 10% 的泛基因组是通过从野生向日葵物种基因渗入得到的,进一步基因功能分析得到这一部分基因与增强抗病性有关。同时研究结果也再一次表明野生近缘种基因的渗入有助于改善栽培品种的抗逆性并增加基因多样性,为未来改善向日葵抗逆性和产油量的研究提供帮助。

2.8 矮牵牛

矮牵牛 (*Petunia hybrida*) 是第一个在其中发现 RNAi (RNA interference) 的物种。商品化的矮牵牛是两种野生矮牵牛亲本的杂交后代,分别为花色白色且由蛾类昆虫授粉的腋叶矮牵牛 (*P. axillaris*) 和花色紫色且由蜜蜂授粉的紫矮牵牛 (*P. inflata*)。Bombarely 等^[31]分别对这两个野生亲本进行了基因组测序,并获得 *P. axillaris* 高质量的基因组序列为 1.26 Gb、*P. inflata* 为 1.29 Gb。这将提高矮牵牛作为一个模式系统的价值,以进一步研究独特的生物现象,比如花色、花香、内源病毒、生物钟、重复序列、微小 RNA、自交不亲和等。

2.9 黄麻

Islam 等^[32]通过对 2 种商用种植黄麻 (*Corchorus olitorius* 和 *C. capsularis*) 组装对比,总共鉴定了 37 031 个 *C. olitorius* 和 30 096 个 *C. capsularis* 的基因。结果表明,两种黄麻基因重复的基因往往偏向于参与对环境刺激的反应。同时,发现了发现 174 个 (53%) *C. olitorius* 和 216 个 (63%) *C. capsularis* 基因在 bre 细胞 [经历次生细胞壁 (SCW) 沉积的细长细胞] 和幼苗中显著表达,揭示了调控纤维生成的

关键基因。

2.10 甘蓝

在农业中甘蓝是一种重要的二倍体植物,如羽衣甘蓝、西兰花等常见园艺作物均包括其中。因为甘蓝基因组存在结构变异:存在-缺失变异 (PAV) 和拷贝数变异 (Copy Number Variation, CNV) 目前无法从参考序列中准确获得所有的基因信息,从而导致其基因组会缺失一些重要的农艺性状基因,造成关联性分析难以进行^[33]。Golicz 等^[34]通过对 9 种甘蓝品种和一种野生型近缘芸薹属物种 *Brassica macrocarpa* 进行基因组测序分析。结果显示甘蓝基因组中,81.3% 为核心基因,非必需基因约占 18.7%。同时发现大量特有 PAV 存在于 *B. macrocarpa* 中,表明在驯化改良过程中甘蓝出现了基因丢失的情况。此外将芸薹属作物与更广泛的物种基因库相比较,发现其等位基因集合有所局限,这可能是由于 PAV 和 CNV 导致的。因此,如果能够与更广泛的物种杂交,可以使芸薹属物种获得一些新的基因,这些新基因的出现有助于增加其物种的多样性和保持杂种优势。

2.11 油菜

Chalhoub 等^[35]对甘蓝型油菜 (*B. napus* cultivar Darmor-bzh) 的纯合子进行了全基因组测序,得到了 8 497 Mb 的基因组草图,通过 SNP (Single Nucleotide Polymorphisms) 遗传图谱分析,将 84% 的基因锚定到了 19 条假定染色体上,其中 10 条亚基因组 An 来自甘蓝基因组,另外 9 条亚基因组 Cn 来源于芸薹基因组。另外发现基因组中转座子 (TEs) 所占比例为 34.8%,且两个亚基因组中 TEs 分布比例不相同。同时通过该研究阐明了甘蓝型油菜基因组 AnCn 的进化历程及在对其选育过程中,能够适应气候、纬度等性状变化的机理进行了解析。Bayer 等^[36]在对两个栽培品种油菜 (*B. napus* cultivar Darmor-bzh 和 *B. napus* cultivar Tapidor) 基因进行 *De Novo* 组装、注释,分别得到 1 345 Mb 和 1 335 Mb。并鉴定了每个品种独有的基因和共同的基因,结果表明 *B. napus* cultivar Darmor-bzh 的基因组与 *B. napus* cultivar Tapidor 相比含有更多的基因,有 73 个基因是后者所没有的,同时也证明了使用共同的注释管道可以导致不同的基因预测,具有重复结构域的基因可能会崩溃,从而低估基因拷贝数。综上研究不仅为甘蓝型油菜基因组的解读提供了新的见解,同时也为芸

藜属物种基因组进化与改良提供有用的资源。

2.12 萝卜

萝卜 (*Raphanus sativus* L.) 同甘蓝一样属十字花科,是主要园艺作物之一,由于其基因组经历了重组、变异等,很难根据现已报道的甘蓝基因组序列进行功能性分析。Kitashiba 等^[37]通过高通量测序获得了一株萝卜 F₂ 自交系(其亲本分别为‘Sayatori 26704’和‘Aokubi S-h’)植株的全基因组草图并预测出 61 572 个基因,其与白菜 *B. rapa* 基因组相近,而比拟南芥 *Arabidopsis thaliana* 基因组大。同时与另外 4 个近缘种(拟南芥 *A. thaliana*、白菜 *B. rapa*、甘蓝 *B. oleracea* 和野萝卜 *R. raphanistrum*) 一起进行基因家族分析,得到萝卜家族基因 24 188 个,其中有 8 759 个为萝卜特有的基因家族,另外有 6 110 个为 5 个物种共有家族基因。萝卜特有基因家族所占比例为 36.2%,远远高于白菜的 15.6% 和拟南芥的 16.2%,表明萝卜具有更多的特有序列及更丰富的多样性,同时也为研究十字花科作物基因的起源、进化、分歧时间、亚基因组之间的互作关系提供参考,另外对挖掘功能基因、改良性状也做出了显著贡献。

2.13 白菜

Conant 等^[38]对中国大白菜 (*B. rapa* accession Chiifu-401-42) 进行基因组测序,注释了约 4 000 个基因,总长为 283.8 Mb。结果表明:反转座子、转座子和长散在重复序列 (Long Interspersed Nuclear Elements, LINE) 分别为 27.1%、3.2% 和 2.8%,大部分存在于着丝粒附近。通过对比拟南芥 *A. thaliana* 发现受转座子 TE 影响,白菜与拟南芥遗传距离较大,白菜基因组存在大量基因丢失情况,其亚基因组 LF、MF1 和 MF2 均有不同程度的丢失。白菜基因组是研究基因组多倍化较好的模型(全基因组三倍化, WGT),该研究为白菜属植物的遗传研究和十字花科作物优良基因选育提供了重要的参考。

2.14 菠菜

Xu 等^[39]通过对 9 个菠菜 (*Spinacia oleracea* L.) 种质进行转录组测序,共生成了大约 1 亿个高质量的碱基对,并将这些碱基对从头组装成 72 151 个单基因,总长度为 46.5 Mb。通过比较这些单基因与不同蛋白质数据库的序列,其中近 60% 被注释,并预测了 387 个代谢途径。同时在栽培种和野生种之间

发现了大量参与生物胁迫和非生物胁迫反应的差异表达基因,这一发现为农业性状改良提供了宝贵资源。

2.15 玉米

玉米作为一种具有高度基因组多态性及杂种优势显著的重要作物,单个个体的参考基因组是不够的,因此泛基因组对提高基因组覆盖度具有重要的意义。简银巧等^[40]对温带玉米自交系 B73 基因组进行补充、优化、注释,同时对 79 份温带玉米自交系重新测序,对其中 31 份热带玉米自交系进行三代转录组混池测序,通过与参考序列注释比对得到 16 121 个已知注释基因、2 991 个未知功能基因和 944 个融合基因。结果表明,热带与温带玉米自交系之间的基因组大小存在明显差异,并发现其基因组大小均与玉米开花显著关联。Sun 等^[41]通过对应雄性代谢株系 Mo17 进行组装得到 2 183 Mb 基因组,其中约 96.4% 序列都集中在 10 个假染色体上,同时有 38 620 个蛋白编码基因被准确注释。并对 B73 和 Mo17 基因组进行比对分析,发现两个基因组之间存在基因结构变异和基因突变,这可能是导致两个自交系之间出现相当大的蛋白质差异原因之一。Springer 等^[42]通过对 W22 基因组进行 *De Novo* 测序组装,结果发现与 B73 基因组相比,W22 存在显著的结构异质性,如转座子组成、拷贝数变异及单核苷酸多样性。以上研究为玉米泛基因组的构建、基因组的演化、功能基因的研究和育种提供了参考。

2.16 杨树

Pinosio 等^[43]通过对 3 个异交杨树全基因组结构变异的分析对比,结果检测到有 3 230 个基因受到 CNV 的影响。7 889 个缺失和 10 586 个插入,覆盖了大约 33.2 Mb 和 62.9 Mb 的基因序列。插入缺失标记 (Insertion-deletion, In Del) 更多的位于低密度基因区,并且与 TE 的活动有关,对于受 In Del 影响的基因进行功能注释,结果表明与抗逆性和抗病性有关的分类较多,同时也说明转座子 TE 的插入对生物性状有显著影响。

2.17 芝麻

Yu 等^[44]组装了 5 个芝麻品种的基因组,得到了一个大小为 554.05 Mb 的芝麻泛基因组。这个泛基因组共有 26 472 个直系同源基因簇,15 409 个核心基因簇 (58.21%),其余的 11 063 个为非核心基

因簇(41.79%)。现代栽培品种中,中国和印度的品种之间存在着显著的基因差异,其特异性基因主要包括与产量、品质相关的基因,而在地方的品种中主要是与环境适应性相关基因。同时揭示了参与植物—病原菌互作、脂类代谢的快速进化和正向选择基因可能与改善芝麻环境适应性和高油脂积累有关。

2.18 小麦

小麦(*Triticum aestivum* L.)作为三大粮食作物之一,其六倍体及庞大的基因组为基因组组装、分子育种工作带来了极大的挑战。Montenegro等^[45]通过对中国小麦品种(Chinese Spring)的基因进行组装测序,并与之前公布的18个小麦的测序数据进行比对,将没有比对上的数据进行过滤、组装和注释,得到了21 653个基因,此外,研究者还将每个品种小麦基因序列与其泛基因组进行比对,发现发现19个小麦都包含的核心基因约为81 070个,同时发现有245个基因是中国小麦品种(Chinese Spring)独有,其它18个小麦中有但在Chinese Spring中没有的基因为12 150个,对非必需基因进行功能富集分析表明其主要与逆境胁迫和防御反应机制有关。国际小麦基因组测序联盟(International Wheat Genome Sequencing Consortium, IWGSC)^[46]报道了一个带有21条染色体的六倍体小麦品种(Chinese Spring)基因组,大小为14.5 Gb。并与3个亚基因组A、B、D的基因及tRNA、TE、microRNA进行比对,发现虽然经历了存在-缺失变异(PAV)后3个亚基因组中TE含量不同,但却具有相似的占比例。同时通过提供一个代表小麦发育主要阶段的转录组图谱,和已知农艺性状质量或数量性状的分析,揭示了参与环境适应和小麦品质形成的复杂基因家族的动态变化,这为加快小麦生物学和基因组学辅助育种的研究和应用提供了新的见解,同时也丰富了可能与重要农艺性状相关的基因,为作物的基因组改良奠定了基础。

2.19 番茄

番茄作为全球消费量最大的蔬菜之一,其泛基因组的构建对未来番茄分子育种、挖掘新的性状起到了至关重要的作用。高磊等^[47]通过对725个栽培番茄和野生番茄的基因进行组装,其中栽培番茄分别为372个SLL(*Solanum lycopersicum* var. *lycopersicum*)及267个SLC(*S. lycopersicum* var. *cerasiforme*);近亲属78个SP(*S. pimpinellifolium*)和8个SCG

(*S. cheesmaniae* and *S. galapagense*)。结果表明有4 873个基因是参考基因组中不存在的,并发现了一种影响水果风味的等位基因TomLoxC,进一步研究发现野生番茄中91.2%的SP有TomLoxC等位基因的存在,但在较老的驯化番茄中SLL只存在2.2%。另外在基因组比对中发现野生番茄具有更多的基因,从另一个角度表明在番茄的驯化和改良过程中,部分基因丢失的趋势。该研究报告不仅为番茄泛基因组构建提供了新的基因,为番茄品质改良带来了新的资源;同时也表明在现代番茄育种中,品质风味特性往往被忽略,而将育种目标较多的集中到对生物及非生物胁迫的抗性、耐储存和高产中,从而导致部分基因丢失,遗传多样性减少。

3 展望

随着测序技术的高速发展,更广泛的物种基因组将会被逐渐测序,人们也意识到单个个体的基因组并不足以代表整个物种基因多样性。因此,对泛基因组学的研究,不仅可以全面地从基因组水平分析物种内遗传多样性,探究个体间的系统发生关系和表型差异的遗传基础^[11,48],而且可以对多个物种、亚种的基因组进行比对分析,挖掘其特有基因和变异位点,为研究物种的起源及演化等重要生物学问题提供依据,并为功能基因研究、科学育种提供指导。此外,通过研究物种泛基因组研究者意识到:某些作物在驯化和改良的过程中,由于过多的将育种目标集中到高产、稳产、抗逆境等方面,一些如品质风味、颜色等观感性状被忽略,使一些相关基因负向选择导致基因的遗失和多样性的减少。最后泛基因组研究还可运用于对不同生态地理类型中差异较大的种质资源进行基因组测序,挖掘物种中新的基因,为候选基因的补充、物种多样性及适应性进化、起源经历和外来物种入侵性等问题的研究提供帮助。

目前,泛基因组研究已在各个领域得到充分的发挥与应用,但是由于庞大的基因信息和测序深度的局限性,即使就目前的基因测序速度而言,也是超出了目前现有的分析能力。因此更新现有技术和方法将给泛基因组学研究带来新的突破,泛基因组学在植物中的研究应用将会成为热点。

参考文献:

[1] Fleischmann R D, Adams M D, White O, et al. Whole-

- genome random sequencing and assembly of *Haemophilus influenzae* Rd[J]. *Science*, 1995, 269(5223):496–512.
- [2] Bult C J, White E O, Olsen G J, et al. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii* [J]. *Science*, 1996, 273(5278):1058–1073.
- [3] Blattner F R, Bloch C A, Perna T, et al. The complete genome sequence of *Escherichia coli* K-12 [J]. *Science*, 1997, 277(5331):1453–1462.
- [4] Tettelin H, Masignani V, Cieslewicz M J, et al. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome” [J]. *Proceedings of the National Academy of Sciences*, 2005, 102(39):13950–13955.
- [5] Hardigan M A, Crisovan E, Hamilton J P, et al. Genome reduction uncovers a large dispensable genome and adaptive role for copy number variation in asexually propagated *Solanum Tuberosum* [J]. *Plant Cell*, 2016, 28(2):388–405.
- [6] 付静, 秦启伟. 30株大肠杆菌的泛基因组学特征分析 [J]. *遗传*, 2012, 34(6):765–772.
Fu J, Qin Q W. Pan-genomics analysis of 30 *Escherichia coli* genomes [J]. *Hereditas*, 2012, 34(6):765–772.
- [7] Tettelin H. The bacterial pan-genome and reverse vaccinology [M]. Basel: Karger Publishers, 2009.
- [8] Schubert-unkmeier A, Christodoulides M. Genome-based bacterial vaccines: current state and future outlook [J]. *BioDrugs*, 2013, 27(5):419–430.
- [9] Tettelin H, Masignani V, Cieslewicz M J, et al. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae* [J]. *Proceedings of the National Academy of Sciences*, 2002, 99(19):12391–12396.
- [10] Li R, Li Y, Zheng H, et al. Building the sequence map of the human pan-genome. [J]. *Nature Biotechnology*, 2010, 28(1):57–63.
- [11] Golicz A A, Batley J, Edwards D. Towards plant pangenomics [J]. *Plant Biotechnology Journal*, 2016, 14(4):1099–1105.
- [12] Li Y, Zhou G, Ma J, et al. De novo assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits [J]. *Nature Biotechnology*, 2014, 32(10):1045–1052.
- [13] Yang X, Li Y, Zang J, et al. Analysis of pan-genome to identify the core genes and essential genes of *Brucella spp* [J]. *Molecular Genetics and Genomics*, 2016, 291(2):905–912.
- [14] Laing C, Buchanan C, Taboada E N, et al. Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions [J]. *BMC Bioinformatics*, 2010, 11(1):461.
- [15] Medini D, Donati C, Tettelin H, et al. The microbial pan-genome [J]. *Current Opinion in Genetics & Development*, 2005, 15(6):589–594.
- [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.
Zhao Y B. Development and application of pan-genomics analysis method [D]. Beijing: Beijing Institute of Genomics, Chinese Academy of Sciences, 2014.
- [17] Lefebvre T, Stanhope M J. Evolution of the core and pan-genome of *Streptococcus*: positive selection, recombination, and genome composition [J]. *Genome Biology*, 2007, 8(5):R71.
- [18] Bentley S. Sequencing the species pan-genome [J]. *Nature Reviews Microbiology*, 2009, 7(3):258–259.
- [19] 杨婷婷. 链霉菌海洋环境适应机制的泛基因组学研究 [D]. 北京: 中国科学院北京基因组研究所, 2015.
Yang T T. Pan-genomics study of marine environmental adaptation mechanisms of *Streptomyces* [D]. Beijing: Beijing Institute of Genomics, Chinese Academy of Sciences, 2015.
- [20] Yin D M, Ji C M, Ma X L, et al. Genome of an allotetraploid wild peanut *Arachis monticola*: a de novo assemble [J]. *Giga-Science*, 2018, 7(6):1–10.
- [21] Bertoli D J, Cannon S B, Froenicke L, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut [J]. *Nature Genetics*, 2016, 48(4):438–446.
- [22] Zhang J, Chen L L, Xing F, et al. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63 [J]. *Proceedings of the National Academy of Sciences*, 2016, 113(35):E5163–E5171.
- [23] Zhao Q, Feng Q, Lu H, et al. Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice [J]. *Nature Genetics*, 2018, 50(2):278–284.
- [24] Stein J C, Yu Y, Copetti D, et al. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza* [J]. *Nature Genetics*, 2018, 50(2):285–296.
- [25] Kim S, Park J, Yeom S I, et al. New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication [J]. *Genome Biology*, 2017, 18(1):210–220.
- [26] Ou L, Li D, Lv J, et al. Pan-genome of cultivated pepper (*Capsicum*) and its use in gene presence-absence variation analyses [J]. *The New Phytologist*, 2018, 220(2):360–363.
- [27] Sun H, Wu S, Zhang G, et al. Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita* genomes [J]. *Molecular Plant*, 2017, 10(10):1293–1306.
- [28] Xu S, Brockmüller T, Navarro-quezada A, et al. Wild

- tobacco genomes reveal the evolution of nicotine biosynthesis[J]. Proceedings of the National Academy of Sciences, 2017, 114(23): 6133–6138.
- [29] Badouin H, Gouzy J, Grassa C J, et al. The sunflower genome provides insights into oil metabolism, flowering and *Asterid* evolution [J]. Nature, 2017, 546 (7656): 148–152.
- [30] Hübner S, Bercovich N, Todesco M, et al. Sunflower pan-genome analysis shows that hybridization altered gene content and disease resistance [J]. Nature Plants, 2019, 5 (1): 54–62.
- [31] Bombarely A, Moser M, Amrad A, et al. Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida* [J]. Nature Plants, 2016, 2(6): 279–289.
- [32] Islam M S, Saito J A, Emdad E M, et al. Comparative genomics of two jute species and insight into fibre biogenesis [J]. Nature Plants, 2017, 3(2): 29–35.
- [33] Hurgobin B, Golicz A A, Bayer P E, et al. Homoeologous exchange is a major cause of gene presence/absence variation in the amphidiploid *Brassica napus* [J]. Plant Biotechnology Journal, 2018, 16(7): 1265–1274.
- [34] Golicz A A, Bayer P E, Barker G C, et al. The pangenome of an agronomically important crop plant *Brassica oleracea* [J]. Nature Communications, 2016(7): 13390.
- [35] Chalhoub B, Denoeud F, Liu S Y, et al. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome [J]. Science, 2014, 345(6199): 950–953.
- [36] Bayer P E, Hurgobin B, Golicz A A, et al. Assembly and comparison of two closely related *Brassica napus* genomes [J]. Plant Biotechnology Journal, 2017, 15 (12): 1602–1610.
- [37] Kitashiba H, Li F, Hirakawa H, et al. Draft sequences of the radish (*Raphanus sativus* L.) genome [J]. DNA Research, 2014, 21(5): 481–490.
- [38] Conant G, Soumpourou E, Hua W, et al. The genome of the mesopolyploid crop species *Brassica rapa* [J]. Nature Genetics, 2011, 43(10): 1035–1039.
- [39] Xu C, Jiao C, Zheng Y, et al. De novo and comparative transcriptome analysis of cultivated and wild spinach [J]. Scientific Reports, 2015, 5(1): 981–997.
- [40] 简银巧. 热带玉米全长泛转录组和基因组大小变异及应用 [D]. 北京: 中国农业科学院, 2017.
- Jian Y Q. Variations in pan-transcriptome and genome size in tropical maize (*Zea mays* L.) and their application [D]. Beijing: Chinese Academy of Agricultural Sciences, 2017.
- [41] Sun S, Zhou Y, Jian C, et al. Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes [J]. Nature Genetics, 2018, 50(9): 1289–1295.
- [42] Springer N M, Anderson S N, Andorf C M, et al. The maize W22 genome provides a foundation for functional genomics and transposon biology [J]. Nature Genetics, 2018, 50(9): 1282–1288.
- [43] Pinosio S, Giacomello S, Faivre-rampant P, et al. Characterization of the poplar pan-genome by genome-wide identification of structural variation [J]. Molecular Biology and Evolution, 2016, 33(10): 2706–2719.
- [44] Yu J, Golicz A A, Lu K, et al. Insight into the evolution and functional characteristics of the pan-genome assembly from sesame landraces and modern cultivars [J]. Plant Biotechnology Journal, 2019, 17(5): 881–892.
- [45] Montenegro J D, Golicz A A, Bayer P E, et al. The pangenome of hexaploid bread wheat [J]. The Plant Journal, 2017, 90(5): 1007–1013.
- [46] Appels R, Eversole K, Feuillet C, et al. Shifting the limits in wheat research and breeding using a fully annotated reference genome [J]. Science, 2018, 361(6403): 7191–7203.
- [47] Gao L, Gonda I, Sun H, et al. The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor [J]. Nature Genetics, 2019, 51(6): 1044–1051.
- [48] 庄绪冉, 朱泳璋. 细菌泛基因组学的研究 [J]. 上海交通大学学报(医学版), 2012, 32(11): 1440–1443.
- Zhuang X R, Zhu Y Z. Research progress of bacterial pan-genome [J]. Journal of Shanghai Jiaotong University (Medical Edition), 2012, 32(11): 1440–1443.

责任编辑: 罗小宁
英文校对: 王芬